



**HAL**  
open science

# Approche quantitative de la réponse écologique des espèces végétales forestières à l'échelle de la France

Christophe Coudun

► **To cite this version:**

Christophe Coudun. Approche quantitative de la réponse écologique des espèces végétales forestières à l'échelle de la France. Sciences du Vivant [q-bio]. Ecole Nationale du Génie Rural des Eaux et Forêts, 2005. Français. NNT: . tel-02833929

**HAL Id: tel-02833929**

**<https://hal.inrae.fr/tel-02833929v1>**

Submitted on 7 Jun 2020

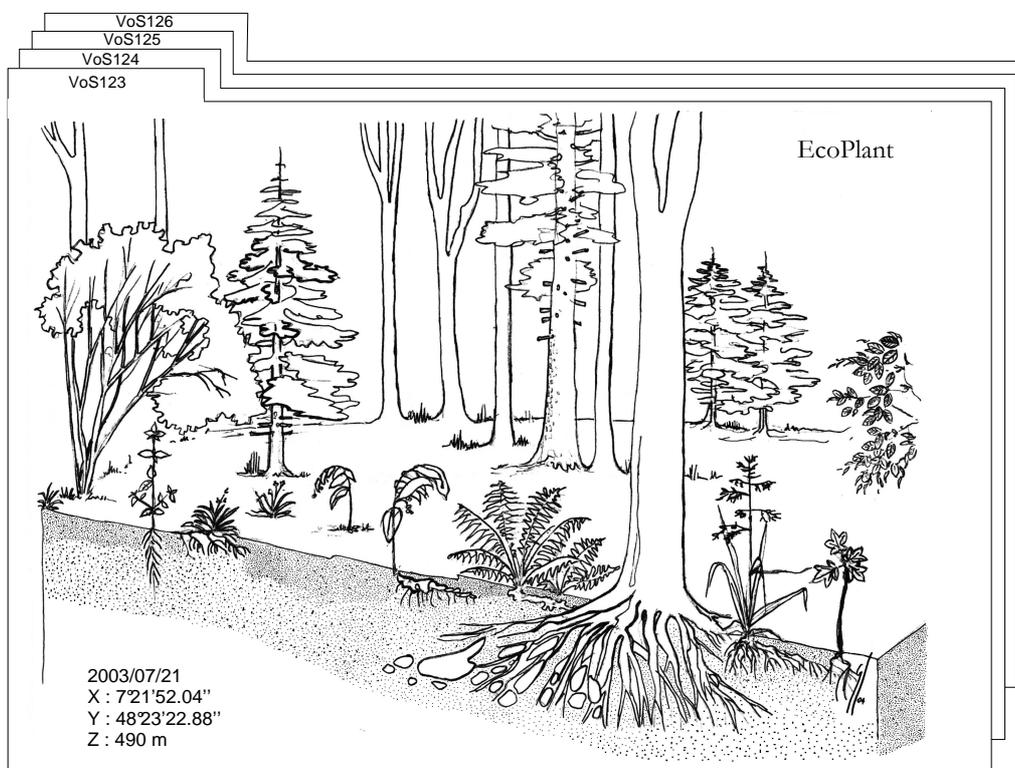
**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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

# Approche quantitative de la réponse écologique des espèces végétales forestières à l'échelle de la France

Thèse de Doctorat en Sciences Forestières  
de l'Ecole Nationale du Génie Rural des Eaux et des Forêts (ENGREF)

Financement : Office National des Forêts et Conseil Régional de Lorraine



**Christophe Coudun**

[coudun@hotmail.com](mailto:coudun@hotmail.com)

Thèse soutenue publiquement à Nancy le 22 juin 2005 devant le jury suivant:

M. Sylvain Plantureux

Mme Sandra Lavorel

M. Martin Hermy

M. Jean-Claude Rameau

M. Jean-Claude Gégout

M. Anthony Lehmann

M. Patrice Mengin-Lecreulx

Professeur, ENSAIA Vandoeuvre

Directrice de Recherche, CNRS Grenoble

Professeur, Katholieke Universiteit Leuven (Belgique)

Professeur, ENGREF Nancy

Maître de Conférence, ENGREF Nancy

Chargé d'étude, CSCF Neuchatel (Suisse)

Chef du Département Recherche, ONF Fontainebleau

Président

Rapporteur

Rapporteur

Directeur

Responsable

Examineur

Examineur



## Remerciements

Merci à **Jean-Claude Gégout** pour sa gentillesse, sa disponibilité et le suivi exceptionnels qu'il m'a accordés depuis plus de cinq ans maintenant... Les félicitations du jury sont également pour lui !

Merci à **Jean-Claude Rameau**, pour son suivi et pour les nombreuses discussions matinales autour du café. Sa disparition quelques mois après ma soutenance laissent un grand vide dans le monde de l'écologie et de la forêt...

Merci à **Christian Piedallu**, pour sa disponibilité et son soutien précieux dans la manipulation des systèmes d'information géographique, et à **Ingrid Seynave**, pour les calculs de valeurs indicatrices dans le dernier chapitre !

Merci à tous les membres de **l'équipe Ecologie Forestière** de l'ENGREF de Nancy, qui m'ont permis d'évoluer dans un cadre de vie et de travail remarquable, et particulièrement à **Isabelle Anselme** et **Evelyne Chaudy** pour leur aide dans la préparation de la soutenance et du pot de thèse, et à **Damien Marage**, « **Pepe** » **Cabello** et **Paulina Pinto** pour avoir partagé la vie de « thésard » en même temps que moi.

Merci à **Laurette Westrich**, pour la reproduction de la thèse.

Merci aux **nombreux auteurs** des études intégrées à EcoPlant, sans qui bien évidemment mon travail de recherche aurait été impossible.

Merci aux membres de mon comité de thèse, et particulièrement à **Jean-Luc Dupouey** et **Jean-Christophe Hervé** pour leurs conseils écologiques ou méthodologiques avisés.

Merci aux membres du jury, et particulièrement à **Sandra Lavorel** et **Martin Hermy** pour leur travail de rapporteur et à **Anthony Lehmann**, pour m'avoir permis de rencontrer mes « auteurs préférés » lors de la semaine Riederalp en août 2004.

Merci aux membres du **Centre d'Etude de la Dynamique du Carbone Terrestre (CTCD)**, et particulièrement à **F. Ian Woodward**, **Shaun Quegan** et **Sam Evans**, pour la confiance qu'ils m'ont accordée en m'offrant un poste trois mois avant la date de soutenance.

Merci à **ma mère**, **mon frère** et **mes sœurs**, pour leur soutien et pour tous les moments passés ensemble.

Merci à **mon père**, parti trop tôt, pour m'avoir toujours soutenu.

Merci à **Julie**, pour tout le reste. Je lui dédie ma thèse...

Farnham, Angleterre, le 27 octobre 2005,

**Christophe Coudun.**



## Résumé

L'objectif de ce travail est de déterminer l'intérêt des bases de données présentant un nombre important de relevés floristiques et écologiques complets pour étudier de façon quantitative le comportement écologique des espèces végétales forestières sur de vastes territoires.

Les données utilisées sont celles d'EcoPlant, développée à l'Ecole Nationale du Génie Rural des Eaux et des Forêts (ENGREF, Nancy) pour stocker des milliers de relevés phytoécologiques complets (relevés floristiques et caractéristiques climatiques et édaphiques précises), réalisés dans les forêts de France. La définition de 54 courbes de réponse théoriques d'espèces végétales vis-à-vis du pH, et la création de jeux artificiels de données binaires de taille variable pour essayer de re-créeer les courbes de réponse théoriques, ont permis de montrer qu'il est difficile de modéliser de manière fiable le comportement des espèces peu fréquentes avec la régression logistique. En effet, nous montrons que 50 à 100 occurrences d'une espèce représentent un seuil minimal pour déterminer précisément l'optimum, l'amplitude et la probabilité maximale de la courbe de réponse écologique des espèces avec la régression logistique, ce qui justifie la nécessité de travailler avec de larges bases de données.

Pour tester la stabilité du comportement écologique des espèces sur de vastes territoires, nous avons comparé la réponse au pH du sol de 46 herbacées forestières entre le nord-est et le nord-ouest de la France, et de 21 herbacées forestières entre les Vosges et le Jura. En contexte de plaine ou de montagne, la majorité des espèces étudiées n'ont pas révélé de différences régionales marquées en terme d'optimum ou d'amplitude écologiques vis-à-vis du pH de l'horizon A du sol. Les quelques différences régionales observées ont été expliquées dans les deux cas par des conditions de compétition différentes dans chaque région. Nous concluons que l'étude de la réponse écologique peut s'effectuer sur de vastes territoires tels que la France, à condition que les conditions de compétition ne soient pas trop différentes selon les régions.

Une caractéristique majeure de la base EcoPlant est de stocker des informations édaphiques et climatiques, couplée à des relevés floristiques effectués sur les mêmes sites. L'étude conjointe de la réponse écologique des espèces végétales vis-à-vis de facteurs climatiques et édaphiques a rarement été menée, puisque la majorité des études étudient la réponse climatique des espèces végétales. Nous illustrons l'importance de prendre en compte des variables édaphiques de nutrition dans les modèles de niche écologique et de distribution géographique des plantes terrestres, par la modélisation de l'écologie et de la distribution de l'Erable champêtre (*Acer campestre* L.) dans les forêts françaises.

L'ensemble des résultats met en évidence l'intérêt des bases de données importantes et écologiquement complètes telles qu'EcoPlant pour approfondir la connaissance des différentes dimensions de la niche écologique des espèces à l'échelle de vastes territoires.

## Mots-clés

Bases de données phytoécologiques, EcoPlant, plantes forestières, niche réalisée, régression logistique, courbes de réponse, optimum écologique, amplitude écologique, simulations numériques, compétition, pH, cartographie prédictive, aire de répartition, *Acer campestre*, France.



## Summary

The objective of this work is to assess the interest of large databases that store a large number of complete floristic and ecological relevés, in order to quantitatively investigate forest plant species ecological behaviour on large areas.

The data used in this work are extracted from EcoPlant, which has been developed at the French Institute of Forestry, Agricultural and Environmental Engineering (ENGREF, Nancy, France), to store thousands of complete phytoecological relevés (floristic relevés with precise climatic and edaphic characteristics), performed in forests from France. The definition of 54 theoretical plant species response curves with regard to pH, and the creation of artificial binary data sets of varying size to try to re-create the theoretical response curves, revealed that it was difficult to model the ecological behaviour of not frequent species in a reasonable way with logistic regression. Indeed, it is shown that 50 to 100 occurrences of a species represent a minimum threshold to accurately determine the optimum, the amplitude, and the maximal probability of the response curve with logistic regression, and this justifies the need to work with large databases.

In order to test the stability of species ecological behaviour on large areas, we compared the response of 46 forest herbs to pH between north-east and north-west France, and the response of 21 forest herbs to pH between the Vosges and the Jura mountains. In either flat or mountainous context, most plant species did not reveal strong regional differences in terms of ecological optimum or amplitude with regard to pH of the first A soil horizon. The slight observed regional differences were explained in both contexts by different competition conditions in each region. We conclude that the investigation of the ecological response can be performed on large areas like France, provided that competition conditions are not too different between studied regions.

A major feature of EcoPlant is to store edaphic and climatic information, coupled with floristic relevés performed at the same sites. The investigation of plant species ecological behaviour to both climatic and edaphic factors has been rarely performed, since most studies currently investigate the climatic behaviour of plant species. We illustrate the need to take edaphic variables into account in ecological niche and predictive distribution models for terrestrial plant species, by modelling *Acer campestre* L. (Field Maple) ecological behaviour and geographic distribution in French forests.

Results show the interest of large databases such as EcoPlant, that are ecologically complete, in order to increase the knowledge of the different dimensions of species ecological niche on large areas.

## Key words

Phytoecological databases, EcoPlant, forest plants, realised niche, logistic regression, response curves, ecological optimum, ecological amplitude, numerical simulations, competition, pH, predictive mapping, geographic range, *Acer campestre*, France.



<b>Résumé</b>	<b>5</b>
Mots-clés	5
<b>Summary</b>	<b>7</b>
Key words	7
<b>Liste des tableaux</b>	<b>12</b>
<b>Liste des figures</b>	<b>12</b>
<b>Introduction</b>	<b>15</b>
 <b>PREMIERE PARTIE : ETUDE DE L'ECOLOGIE DES ESPECES AVEC DE GROSSES BASES DE DONNEES</b>	
 <b>1. EcoPlant : la base de données phytoécologiques forestières développée par l'ENGREF pour lier des données floristiques à des variables du sol et du climat</b>	
	<b>23</b>
<b>Résumé</b>	<b>24</b>
Mots-clés	24
<b>1.1. Introduction</b>	<b>25</b>
<b>1.2. Quel type d'information stocker dans EcoPlant ?</b>	<b>26</b>
<b>1.3. Quelles variables pertinentes stocker dans EcoPlant ?</b>	<b>27</b>
1.3.1. Des caractéristiques complètes des sources des études	27
1.3.2. Des informations précises sur la localisation des relevés	27
1.3.3. Des relevés floristiques dans une nomenclature cohérente et unifiée	28
1.3.4. Des données climatiques issues de croisements sous SIG	28
1.3.5. Des données édaphiques issues de mesures sur le terrain et d'analyses en laboratoire	29
<b>1.4. Chiffres clés d'EcoPlant</b>	<b>29</b>
1.4.1. 200 champs floristiques et écologiques	29
1.4.2. 120 sources de relevés et 6 432 relevés phytoécologiques forestiers	30
1.4.3. 2 154 taxons	32
<b>1.5. Conclusions et perspectives</b>	<b>34</b>
1.5.1. Gestion des données et contrôle de la qualité	34
1.5.2. Applications actuelles d'EcoPlant	34
1.5.3. Perspectives	35
 <b>2. Modélisation de la niche écologique des espèces végétales par la régression logistique et transposition des modèles dans l'espace géographique</b>	
	<b>37</b>
<b>Résumé</b>	<b>38</b>
Mots-clés	38
<b>2.1. Introduction</b>	<b>39</b>
<b>2.2. Définition du modèle de régression logistique</b>	<b>39</b>

<b>2.3. Interprétation des modèles dans l'espace écologique : les courbes et surfaces de réponse</b>	<b>41</b>
2.3.1. Courbes et surfaces de réponse écologique	41
2.3.2. Optimum et amplitude écologiques	42
<b>2.4. Interprétation des modèles dans l'espace géographique : les cartes de distribution potentielle</b>	<b>43</b>
2.4.1. Spatialisations des modèles	43
2.4.2. Evaluation des cartes de distribution potentielle	43
<b>2.5. Discussion</b>	<b>45</b>
<b>3. L'établissement de courbes de réponse écologique avec la régression logistique est sensible à l'intensité de l'échantillonnage et aux caractéristiques des courbes</b>	<b>47</b>
<b>Résumé</b>	<b>48</b>
Mots-clés	48
<b>3.1. Introduction</b>	<b>49</b>
<b>3.2. Matériels et méthodes</b>	<b>51</b>
3.2.1. Régression logistique et résumés numériques des courbes de réponse	51
3.2.2. Courbes de réponse théoriques	51
3.2.3. Jeux de données artificiels de taille variable	52
3.2.4. Calcul et évaluation de la qualité des modèles	53
3.2.5. Application à des données réelles	53
<b>3.3. Résultats</b>	<b>54</b>
3.3.1. Prédiction de la réponse des espèces au facteur écologique	54
3.3.2. Le nombre d'occurrences est le critère qui permet de sélectionner les espèces à étudier	55
3.3.3. Qualité des modèles calculés et caractéristiques des courbes de réponse	56
3.3.4. Application à des données réelles	59
<b>3.4. Discussion</b>	<b>61</b>

## DEUXIEME PARTIE : STABILITE DU COMPORTEMENT DES ESPECES A L'ECHELLE DE LA FRANCE

<b>4. Comparaison de la réponse écologique des espèces forestières herbacées vis-à-vis du pH du sol dans le nord-est et le nord-ouest de la France</b>	<b>65</b>
<b>Résumé</b>	<b>66</b>
Mots-clés	66
Nomenclature	66
<b>4.1. Introduction</b>	<b>67</b>
<b>4.2. Matériels et méthodes</b>	<b>68</b>
4.2.1. Deux jeux de données régionaux	68
4.2.2. Modélisation par régression logistique simple	69
<b>4.3. Résultats</b>	<b>70</b>
4.3.1. Stabilité générale de la réponse au pH du sol pour la plupart des espèces	70
4.3.2. La compétition pour les environnements les plus favorables	71
<b>4.4. Discussion et conclusions</b>	<b>74</b>

## 5. Comparaison des pools d'espèces, de la richesse spécifique et de la réponse écologique des espèces forestières herbacées vis-à-vis du pH du sol dans les Vosges et le Jura 77

<b>Résumé</b>	<b>78</b>
Mots clés	78
<b>5.1. Introduction</b>	<b>79</b>
<b>5.2. Matériels et méthodes</b>	<b>80</b>
5.2.1. Régions d'étude et données disponibles	80
5.2.2. Caractérisation de la composition floristique	81
5.2.3. Caractérisation du comportement écologique	83
<b>5.3. Résultats</b>	<b>83</b>
5.3.1. Composition floristique dans les Vosges et le Jura	83
5.3.2. Réponse des espèces au pH	86
<b>5.4. Discussion</b>	<b>88</b>

## TROISIEME PARTIE : ETUDE DE LA NICHE ECOLOGIQUE DES ESPECES A L'ECHELLE DE LA FRANCE

### 6. Les facteurs nutritionnels du sol améliorent les modèles de prédiction de la distribution des espèces végétales : une illustration avec l'Erable champêtre (*Acer campestre* L.) en France 93

<b>Résumé</b>	<b>94</b>
Mots-clés	94
<b>6.1. Introduction</b>	<b>95</b>
<b>6.2. Matériels et méthodes</b>	<b>96</b>
6.2.1. Jeu de données de calibration	96
6.2.2. Jeu de données de validation	97
6.2.3. Régression logistique	98
6.2.4. Evaluation quantitative des modèles	99
6.2.5. Autocorrélation spatiale	100
<b>6.3. Résultats</b>	<b>101</b>
6.3.1. Réponse climatique d' <i>Acer campestre</i>	101
6.3.2. Réponse climatique et édaphique d' <i>Acer campestre</i>	103
<b>6.4. Discussion</b>	<b>107</b>
6.4.1. Remarques méthodologiques	107
6.4.2. Autocorrélation spatiale	107
6.4.3. L'importance des variables nutritionnelles du sol	108
<b>Conclusions et perspectives</b>	<b>109</b>
<b>Références bibliographiques</b>	<b>115</b>

## Liste des tableaux

Tableau 1.1. Chiffres principaux d'EcoPlant, calculés pour chaque type de source de données.....	31
Tableau 2.2. Définition de la table de contingence entre présence/absence observée et présence/absence prédite. .	44
Tableau 3.3. Liste de 21 études de modélisation de la distribution des espèces .....	50
Tableau 3.4. Taille nécessaire du jeu de données pour obtenir 95 % de bons modèles.....	58
Tableau 4.5. Liste des espèces forestières herbacées qui présentent un changement d'optimum ou d'amplitude de pH supérieur à 0,5 unité entre les deux régions d'étude NW et NE.....	73
Tableau 6.6. Liste des variables climatiques édaphiques disponibles pour modéliser la distribution d' <i>Acer campestre</i> dans les forêts de France.....	99
Tableau 6.7. Résumé des modèles calculés par régression logistique.....	102
Tableau 6.8. Pouvoir prédictif des modèles de distribution d' <i>Acer campestre</i> .....	105

## Liste des figures

Figure 1.1. Vue semi-réaliste d'un relevé particulier d'EcoPlant, avec ses caractéristiques de végétation et de sol. ....	26
Figure 1.2. Structure simplifiée d'EcoPlant.....	30
Figure 1.3. Localisation géographique des 6 432 relevés forestiers intégrés à EcoPlant. ....	31
Figure 1.4. Conditions écologiques des placettes intégrées à EcoPlant.....	32
Figure 1.5. Espèces les plus fréquentes d'arbres, d'arbustes, d'herbes et de mousses intégrées dans EcoPlant. ....	33
Figure 1.6. Fréquence cumulée des espèces végétales forestières en France, basée sur 6 432 relevés floristiques et 2 154 taxons, rangés par ordre de fréquence décroissante. ....	33
Figure 1.7. Courbes de réponse de <i>Vaccinium myrtillus</i> , <i>Anemone nemorosa</i> et <i>Cornus mas</i> le long du gradient pH, dans les forêts de France.....	34
Figure 2.8. Exemple schématique d'une courbe de réponse le long d'un gradient écologique.....	41
Figure 2.9. Exemple schématique d'une surface de réponse le long de deux gradients écologiques. ....	42
Figure 2.10. Courbe de réponse, optimum et amplitude écologiques de <i>Crataegus laevigata</i> et <i>Crataegus monogyna</i> vis-à-vis de la température moyenne annuelle.....	42
Figure 2.11. Carte schématique des habitats favorables pour une espèce donnée pour deux variables considérées séparément ou simultanément. ....	43
Figure 2.12. Exemple de courbe ROC. ....	44
Figure 3.13. Probabilité de présence théorique des espèces simulées le long du gradient de pH.....	52
Figure 3.14. Succès de prédiction avec une taille d'échantillon variable, du taux de réaction ou de non-réaction des espèces.....	54
Figure 3.15. Courbes de réponse simulées pour la courbe théorique avec $opt = 4$ , $tol = 1,5$ et $pmax = 0,20$ , et avec une taille variable du jeu de données. ....	55
Figure 3.16. Différence entre optimum calculé et optimum théorique en fonction de la taille du jeu de données, du nombre d'occurrences et de la prévalence. ....	56
Figure 3.17. Qualité des courbes de réponse simulées pour des espèces à faible occurrence. ....	57
Figure 3.18. Nombre de bons modèles simulés pour deux tailles de jeux de données (100 relevés et 1 000 relevés). ....	57
Figure 3.19. Fréquence décroissante des 400 plantes forestières les plus fréquentes intégrées à EcoPlant. ....	59
Figure 3.20. Optimum et amplitude écologiques des 50 plantes les plus fréquentes présentes dans EcoPlant, calculés sur la base de 3 810 relevés et 100 relevés. ....	60
Figure 4.21. Localisation des 2 402 relevés forestiers étudiés dans les deux régions du nord de la France (nord-ouest et nord-est).....	69
Figure 4.22. Courbes de réponse vis-à-vis du pH dans le NW et le NE pour <i>Deschampsia flexuosa</i> , <i>Melica uniflora</i> et <i>Carex flacca</i> . ....	70
Figure 4.23. Représentation des positions des optima et des amplitudes de pH dans le NW et le NE pour les 46 espèces forestières herbacées. ....	71
Figure 4.24. Teneur moyenne du premier horizon minéral du sol (horizon A) en aluminium échangeable, en somme de cations nutritifs échangeables, et en calcaire actif le long du gradient de pH, dans le nord de la France. ....	72
Figure 4.25. Nombres moyens d'espèces herbacées dans le NW et le NE le long du gradient de pH. ....	73

Figure 5.26. Distribution spatiale des placettes comportant un relevé floristique et une mesure de pH du premier horizon du sol dans les Vosges (745 relevés) et le Jura (205 relevés).....	80
Figure 5.27. Distribution des pH des relevés disponibles dans les Vosges (745 relevés) et le Jura (205 relevés).....	81
Figure 5.28. Taille du pool d'espèces herbacées dans les Vosges et le Jura, en fonction du nombre de relevés floristiques.....	84
Figure 5.29. Taille du pool des espèces herbacées dans les Vosges et le Jura, le long du gradient pH. ....	84
Figure 5.30. Indices moyens de similarité de Dice des relevés floristiques le long du gradient de pH, pour les relevés des Vosges, du Jura et de l'ensemble des deux massifs .....	85
Figure 5.31. Nombre d'espèces forestières herbacées par relevé le long du gradient pH. ....	86
Figure 5.32. Exemples de courbes de réponse par rapport au pH, de <i>Dryopteris carthusiana</i> et de <i>Carex sylvatica</i> , dans les Vosges et le Jura. ....	87
Figure 5.33. Histogramme des optima des espèces présentes plus de 10 fois et présentant une réaction significative au pH dans les Vosges et le Jura. ....	87
Figure 5.34. Optimum de pH dans les Vosges et le Jura pour 21 espèces forestières herbacées présentant une réaction significative au facteur dans les deux massifs. ....	88
Figure 6.35. Localisation des 3 286 relevés forestiers étudiés en France (jeu de données de calibration), avec l'indication de la présence et de l'absence d' <i>Acer campestre</i> .....	96
Figure 6.36. Semivariogramme calculé avec la technique « indicator kriging » sur les données de présence/absence d' <i>Acer campestre</i> , pour le jeu de données d'évaluation.....	100
Figure 6.37. Cartes des jeux d'évaluation successifs avec une distance croissante entre les points d'évaluation et entre les points d'évaluation et les points de calibration. ....	101
Figure 6.38. Surface de réponse climatique d' <i>Acer campestre</i> , vis-à-vis des précipitations moyennes automnales et de l'évapotranspiration réelle selon Thornthwaite, calculée avec 3 286 relevés forestiers. ....	103
Figure 6.39. Courbe de réponse écologique d' <i>Acer campestre</i> vis-à-vis du pH du sol, calculé par régression logistique sur 3 286 relevés forestiers.....	104
Figure 6.40. Courbes ROC associées au modèle climatique et au modèle climatique et édaphique. ....	104
Figure 6.41. Cartes de la présence/absence réelle d' <i>Acer campestre</i> et des probabilités de présence prédites par le modèle climatique et le modèle climatique et édaphique.....	106



## Introduction



Deux visions théoriques de la réponse de la végétation aux facteurs du milieu se sont confrontées au cours de l'histoire récente de l'écologie végétale (Goodall 1963, Whittaker 1967, Werger et van der Maarel 1978), avec d'une part la vision de la végétation comme un continuum, faisant l'hypothèse d'une variation continue de la composition floristique le long des gradients écologiques (Gleason 1917, 1926), et d'autre part la vision de la végétation divisée en communautés qui ont évolué ensemble et qui forment des unités homogènes, bien démarquées et reconnaissables (Clements 1936). Un débat pour départager ces deux théories est mené depuis le milieu des années 1960 (Whittaker 1967), mais c'est la théorie du continuum qui s'est imposé au fil du temps (Austin et Smith 1989). Un argument majeur pour étudier la réponse des espèces plutôt que celles des communautés est l'évidence paléoécologique que les communautés d'espèces n'ont jamais été stables au cours du temps, ne se déplaçant probablement pas comme une seule entité avec les changements climatiques (Guisan et Zimmermann 2000).

Pour comprendre et formaliser la relation entre une espèce particulière et son environnement, le concept de niche écologique, développé à l'origine pour les espèces animales, a été adapté pour les espèces végétales (Daget et Godron 1982, Austin 1985). Analyser la niche écologique des espèces permet en effet de décrire et d'analyser la manière dont les espèces interagissent entre elles et vis-à-vis des ressources disponibles (Chase et Leibold 2003). Hutchinson (1957) a décrit la niche d'une espèce comme une région (un hyper-volume à  $n$  dimensions) au sein d'un espace environnemental multidimensionnel composé par des facteurs écologiques (biotiques et abiotiques). Chaque gradient écologique qui a une influence sur la survie de l'espèce est ainsi représenté par un axe et la niche peut se décomposer sur  $n$  axes écologiques. Hutchinson (1957) a distingué la niche fondamentale de la niche écologique (ou niche réalisée), car la première n'est fonction que de la performance physiologique de l'espèce et des contraintes environnementales, tandis que la seconde inclut les interactions biotiques. Dans la nature, c'est la niche écologique (ou niche réalisée) que l'on peut observer.

Dans le monde, l'étude de l'écologie des plantes a concerné principalement la relation entre les espèces et le climat (par exemple Woodward et Williams 1987, Box et al. 1993, Box 1996), puisque le climat est le premier facteur, à l'échelle mondiale, ayant une influence sur la distribution des espèces végétales. En Europe, l'écologie des plantes est étudiée depuis longtemps et des travaux majeurs ont été menés en Europe centrale par Ellenberg et al. (1992). Les travaux d'Ellenberg et al. ont permis de définir un catalogue de valeurs indicatrices empiriques pour des gradients édaphiques et climatiques, en ordonnant les espèces selon leurs préférences écologiques. Sources d'information sur les préférences des plantes vis-à-vis notamment de facteurs édaphiques, les valeurs indicatrices d'Ellenberg ont été abondamment utilisées pour étudier par exemple les niveaux d'acidité et de nutrition en azote des sols (Diekmann 1996, Hawkes et al. 1997, Hill et Carey 1997, Ertsen et al. 1998), ou les changements à long terme des caractéristiques édaphiques des sols (Thimonier et al. 1994, Diekmann et Dupré 1997).

En France, il n'existe qu'un petit nombre de travaux d'autécologie des plantes qui concernent la totalité du territoire. Par exemple, le comportement écologique des plantes vis-à-vis du climat a été caractérisé de manière quantitative grâce aux travaux notamment de Brisse et al. (Brisse et Grandjouan 1980, Garbolino 2001). Dans la Flore Forestière Française (Rameau et al. 1989, 1993), le comportement écologique des plantes forestières vis-à-vis de caractéristiques édaphiques et hydriques a été déterminé de manière qualitative. A l'échelle locale, le comportement écologique des plantes forestières a été déterminé de manière qualitative dans les catalogues de stations forestières (Brêthes 1989). La majorité des études visant à caractériser l'écologie des

espèces forestières (les catalogues de stations forestières) ont donc été menées au niveau local de la petite région naturelle, et avec des nombres de relevés variables et des descripteurs écologiques relevés sur le terrain.

Les études locales et de rares études menées sur de plus vastes régions semblent montrer des variations de l'écologie des espèces d'une région à l'autre. Certes les valeurs indicatrices d'Ellenberg *et al.* (1992) ont été utilisées avec succès dans d'autres régions que celle pour laquelle elles avaient été définies (Ertsen *et al.* 1998, Schaffers et Sykora 2000), mais elle ont dû être adaptées aux conditions locales de certains pays (par exemple Hill *et al.* 1999, 2000 pour la Grande-Bretagne). Certains pays ont même développé leur propre système de valeurs indicatrices (voir par exemple Noirfalise et Dethioux 1970 pour la Belgique ou Landolt 1977 pour la Suisse). En France, Le Tacon et Timbal (1972, 1973) ont montré des différences de comportement écologique des plantes forestières vis-à-vis de l'humus entre le nord-est et le nord-ouest.

Pour comprendre les changements de comportement écologique d'une région à une autre, trois mécanismes principaux ont été proposés (Diekmann et Lawesson 1999) : (i) un changement des besoins physiologiques d'une espèce peut conduire à la différenciation génétique d'écotypes, qui peuvent présenter des comportements différents dans des régions différentes ; (ii) des conditions biotiques différentes peuvent causer des conditions de compétition différentes ; (iii) des conditions écologiques différentes peuvent forcer une espèce à déplacer son habitat pour retrouver des conditions favorables (phénomène de compensation de facteurs, Walter et Walter 1953)

Pour évaluer la niche écologique d'une espèce de manière quantitative, on étudie traditionnellement la réponse écologique des espèces par rapport à des variables environnementales bien définies et continues. On a souvent accès à des données binaires de présence/absence des espèces (on note 1 la présence de l'espèce et 0 son absence), associées à des données de milieu relevées sur les mêmes placettes. On peut ainsi traduire la réponse d'une espèce à un facteur par l'établissement d'une courbe de réponse, qui représente la probabilité de présence de l'espèce pour toutes les valeurs du facteur. Dans les années 1980, le développement des modèles linéaires généralisés (GLMs) a permis les premières études de caractérisation des courbes de réponses des espèces aux facteurs du milieu (Austin *et al.* 1983, 1984, ter Braak et Looman 1986, McCullagh et Nelder 1997). Depuis, de nombreuses techniques statistiques alternatives ont été mises en œuvre, parmi lesquelles on trouve les modèles additifs généralisés (GAMs, Yee et Mitchell 1991, Hastie et Tibshirani 1997, Bio *et al.* 1998) qui sont une extension des GLMs, ou les arbres de classification et de régression (Vayssières *et al.* 2000, Thuiller *et al.* 2003a). Certaines études comparent différentes techniques de modélisation des relations espèces/environnement sur le même jeu de données (Manel *et al.* 1999a, 1999b, Segurado et Araújo 2004). Parmi toutes les techniques de caractérisation directe des relations entre la présence/absence d'une espèce et les conditions environnementales, la régression logistique est actuellement la technique la plus largement utilisée (ter Braak et Looman 1986, McCullagh et Nelder 1997), puisque c'est une technique paramétrique (on peut définir la courbe de réponse par une équation), flexible (on peut étudier la réponse des espèces à des variables qualitatives et quantitatives séparément ou simultanément), et facilement mise en œuvre avec des logiciels de statistique.

Le développement des outils informatiques de calcul et de stockage a pu, en parallèle du développement des techniques statistiques, permettre à de nombreux écologues de traiter des données de manière plus rapide. L'accès à de larges bases de données et à des logiciels de traitement de données adaptés a récemment permis le traitement de plusieurs milliers, voire dizaines de milliers de placettes simultanément pour établir le comportement des espèces étudiées. De nombreuses bases de données phytosociologiques ont été ainsi créées ces dernières années un peu partout dans le monde, afin de stocker de nombreux relevés floristiques accumulés au cours du temps (voir par exemple [Mucina et al. 1993](#), [Lawesson et al. 1997](#), [Ewald 2001](#), [Wiser et al. 2001](#), [Chytry et Rafajova 2003](#)). En France, la base de données Sophy ([Brisse et al. 1995](#)) a été créée pour stocker les relevés floristiques issus des études phytosociologiques. Son contenu purement floristique limite cependant les possibilités d'analyse écologique du comportement des espèces.

La réalisation de nombreux catalogues de stations forestières durant les années 1980-1990 dans toute la France a conduit à la réalisation de centaines de relevés floristiques sur des sites bien décrits écologiquement notamment grâce à des analyses physiques et chimiques de sol réalisées en laboratoire. L'École Nationale du Génie Rural des Eaux et des Forêts (ENGREF) a développé la base de données EcoPlant pour recueillir ces centaines de relevés avec analyses de sol (les exemples caractéristiques des types de stations forestières) et pallier ainsi à leur dispersion dans une centaine de catalogues de stations forestières ([Gégout 2001](#), [Gégout et al. 2005](#)). La finalité d'EcoPlant est de permettre l'étude conjointe de la réponse des espèces végétales forestières à des facteurs du sol et du climat, sur la base de nombreux relevés provenant de diverses sources et répartis dans l'ensemble des écosystèmes forestiers français. La constitution de la base EcoPlant a été finalisée en 2002 et elle contient à ce jour 6 432 relevés floristiques et écologiques complets provenant de 120 études différentes. Des analyses de sol réalisées en laboratoire sont disponibles pour environ 4 000 placettes et des données climatiques obtenues par traitements sous système d'information géographique (SIG) sont disponibles pour 5 500 placettes.

En s'appuyant sur des outils d'analyse de données développés depuis quelques dizaines d'années (la régression logistique notamment), l'objectif principal de la thèse est, à partir de la base de données EcoPlant, d'explorer les possibilités ouvertes par les grosses bases de données phytoécologiques concernant de vastes territoires, pour étudier le comportement écologique des espèces végétales forestières. Trois voies distinctes ont été explorées pour essayer de répondre aux trois questions suivantes et chacune de ces voies fera l'objet d'une partie de la thèse:

- (i) Y-a-t'il un intérêt à travailler avec un nombre important de relevés lorsque l'on veut caractériser le comportement écologique des espèces végétales avec la régression logistique ?**
- (ii) Peut-on caractériser le comportement écologique des espèces végétales forestières à l'échelle de la France ou doit-on régionaliser ce comportement ?**
- (iii) Dans quelle mesure l'apport des variables du sol en complément des variables du climat permet d'appréhender de manière plus complète le comportement écologique des espèces végétales forestières ?**

Pour tenter d'apporter des réponses à ces trois questions, nous présentons dans une première partie les matériels et méthodes de la thèse, à savoir la base de données phytoécologiques EcoPlant, qui fait l'objet du premier chapitre, puis la technique de régression logistique, qui fait l'objet du second chapitre. Dans le troisième chapitre, nous souhaitons évaluer le nombre de relevés nécessaires pour calculer la courbe de réponse des espèces de manière fiable avec la régression logistique et déterminer ainsi leur optimum et leur amplitude écologiques.

Dans la seconde partie de la thèse, nous testons la validité spatiale du comportement écologique des espèces végétales forestières sur de larges territoires, en comparant la réponse d'espèces herbacées dans des régions différentes, d'abord en contexte de plaine, puis en contexte de montagne. La comparaison de la réponse de 46 espèces herbacées forestières au pH de l'horizon A du sol entre le nord-est et le nord-ouest de la France constitue le quatrième chapitre de la thèse, et la comparaison de la réponse de 21 espèces herbacées forestières au pH de l'horizon A du sol entre les Vosges et le Jura, en constitue le cinquième chapitre.

Dans la troisième partie (sixième chapitre) nous caractérisons la réponse écologique de l'Erable champêtre (*Acer campestre* L.) à des variables climatiques et édaphiques en France, puis nous spatialisons les probabilités de présence issues des modèles de régression logistique. Comparée à la carte de distribution réelle de l'Erable champêtre, sa carte de probabilités de présence est analysée pour mieux mettre en relief les effets respectifs du climat et du sol sur la distribution de cette espèce en France.

## **Première partie**

Etude de l'écologie des espèces avec de grosses bases de données



**1. EcoPlant : la base de données phytoécologiques forestières développée par l'ENGREF pour lier des données floristiques à des variables du sol et du climat**

*Article associé (Annexe 1)*

**Gégout, J.-C., Ch. Coudun, G. Bailly et B. Jabiol (2005)** EcoPlant: a forest site database to link floristic data with soil resources and climatic conditions. *Journal of Vegetation Science*, **16**, 257-260.

## Résumé

Afin de pouvoir stocker des milliers de relevés phytoécologiques complets, réalisés dans des forêts de tout le territoire de la France métropolitaine et de la Corse, une base de données phytoécologiques, EcoPlant, a été développée à l'Ecole Nationale du Génie Rural, des Eaux et des Forêts (ENGREF). Le caractère original d'EcoPlant est de stocker des informations floristiques, climatiques et édaphiques complètes, issues d'observations réalisées sur les mêmes sites, avec l'accent mis sur de l'information pédologique très précise (descriptions des profils de sols et des horizons, analyses physiques et chimiques d'échantillons de sol). Le développement d'une telle base de données illustre la volonté du secteur forestier de réaliser, depuis les années 1980, la typologie des stations forestières françaises, en couplant des informations floristiques et édaphiques relevées sur les mêmes sites. Actuellement, 6 432 relevés de grande valeur floristique et écologique sont intégrés à EcoPlant, qui est complémentaire des autres bases de données floristiques ou pédologiques existantes. L'accès à des variables climatiques directes est possible à travers la modélisation, par systèmes d'information géographique (SIG), de données de précipitations et de températures spatialisées par Météo France (modèle Aurelhy). De plus, la disponibilité de facteurs du sol peut permettre une recherche approfondie sur la modélisation de la niche écologique des espèces végétales forestières et sur la prédiction de leur distribution géographique.

## Mots-clés

Base de données phytoécologiques, variables de nutrition, analyses physiques et chimiques du sol, variables climatiques, Systèmes d'Information Géographique, espèces végétales forestières, relevé phytoécologique, France, European Vegetation Survey, relations espèces-environnement.

## 1.1. Introduction

Les études concernant les relations entre les espèces et leur environnement constituent une littérature fondamentale de l'écologie végétale. Basées sur des observations de terrain, elles sont généralement fondées sur le lien entre l'information écologique et floristique relevée sur les mêmes sites. L'analyse conjointe de données floristiques et écologiques a plus particulièrement consisté, pendant ces dernières décennies, en l'établissement de courbes de réponse écologiques, basées sur le concept central de niche écologique et sur ses procédures de modélisation (Guisan et Zimmermann 2000, Guisan et al. 2002, Lehmann et al. 2002a, Scott et al. 2002), avec des applications, par exemple, à la prédiction et la cartographie de la distribution des espèces ou des communautés dans un contexte de changements globaux (Guisan et al. 1998, Guisan et Theurillat 2000a, 2000b).

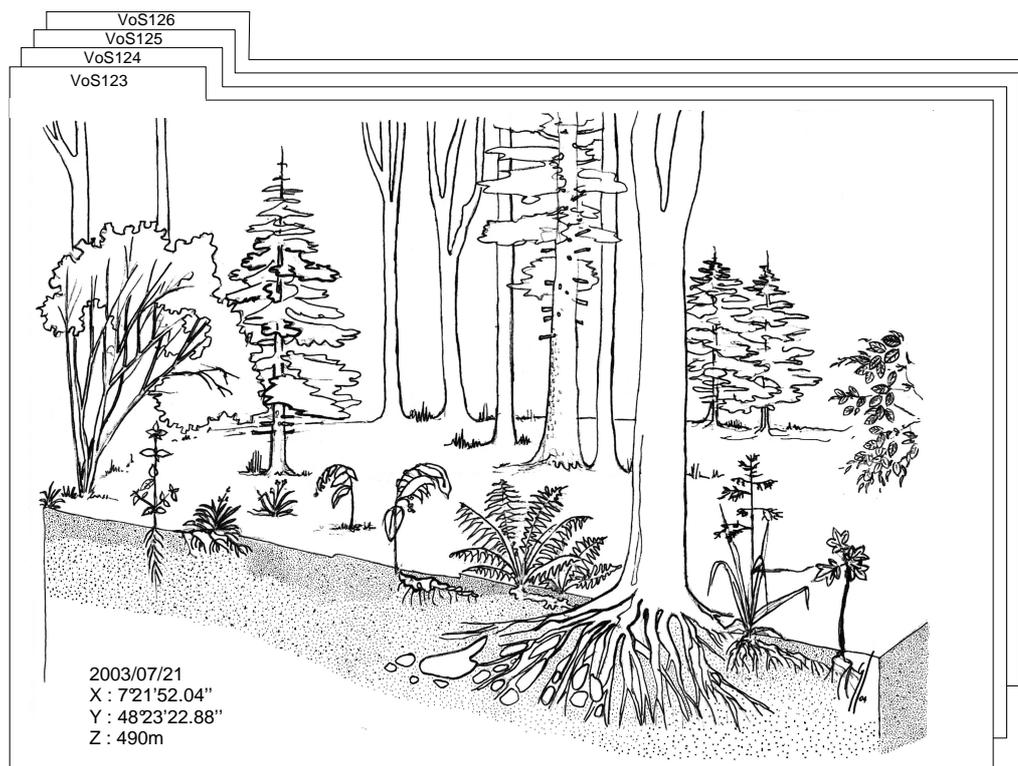
Traditionnellement, les jeux de données phytoécologiques utilisés dans la plupart des études ne concernent qu'un nombre restreint de relevés, dans un contexte écologique et géographique limité. Néanmoins, le développement récent de larges bases de données permet l'analyse du comportement écologique des espèces ou des communautés sur de larges territoires et représente ainsi une voie intéressante pour améliorer notre connaissance de la végétation (Peet et Wiser 2004). Les bases de données de végétation existantes consistent la plupart du temps en des bases de données phytosociologiques à une large échelle (Mucina et al. 1993, Brisse et al. 1995, Lawesson et al. 1997, Mucina et al. 2000a, Ewald 2001, Wiser et al. 2001, Chytrý et Rafajová 2003), en des jeux de données stockés dans des systèmes de gestion tels que TURBOVEG (Hennekens et Schaminée 2001), JUICE (Tichý 2002) ou FLORA-sys (Plantureux 1996), ainsi qu'en des bases de données de traits de vie (Bakker et al. 2000, Jäger 2000, Knevel et al. 2003, Poschlod et al. 2003). En utilisant l'information contenue dans ces bases de données, les relations espèces-environnement sont souvent examinées par des analyses de gradients indirectes (par exemple des analyses factorielles des correspondances), l'autécologie des espèces étant ensuite caractérisée grâce à des variables écologiques indirectes (n'ayant pas d'influence physiologique directe sur les plantes), telles que l'altitude, l'exposition, la pente ou la topographie (par exemple Austin et al. 1983) ou à des valeurs indicatrices écologiques empiriques comme celles d'Ellenberg (Ellenberg et al. 1992, Diekmann 1996, Hawkes et al. 1997). L'utilisation de gros jeux de données floristiques, combinés à des couches écologiques spatialisées dans un système d'information géographique (SIG) est une voie récente pour lier directement l'information floristique observée aux variables climatiques prédictives (Austin et al. 1990, Franklin 1998, Guisan et al. 1998). Cette approche peut s'appliquer aux variables qui peuvent être cartographiées, telles que les variables climatiques qui varient généralement à une large échelle, mais elle ne peut s'appliquer à des variables qui sont difficiles à modéliser spatialement, comme les caractéristiques nutritionnelles du sol.

Nous présentons dans ce chapitre EcoPlant, une base relationnelle de données phytoécologiques qui a été conçue pour permettre la caractérisation quantitative des relations entre la distribution des espèces et leur environnement, sur la base de mesures réelles (Gégout 2001, Gégout et al. 2005). EcoPlant permet le stockage de relevés phytoécologiques contenant, pour chaque site : (i) un relevé floristique exhaustif ; (ii) des variables écologiques mesurées pertinentes qui ont un impact physiologique direct ou indirect sur la croissance et la survie des espèces, (iii) des informations standardisées sur la manière dont le relevé phytoécologique a été réalisé, comme celles proposées par Mucina et al. (2000b).

La possibilité de modéliser la niche écologique des espèces avec à la fois des variables de ressources édaphiques et des variables climatiques qui ont un impact direct sur la distribution des plantes, constitue ainsi la principale motivation de la constitution d'EcoPlant. Le choix délibéré d'intégrer parmi ces variables, celles qui sont sensibles aux changements globaux en termes de température, d'acidité ou de conditions d'azote, doit permettre une meilleure compréhension des écosystèmes forestiers dans leur évolution à long terme.

## 1.2. Quel type d'information stocker dans EcoPlant ?

Chaque relevé stocké dans EcoPlant a été effectué à un moment donné et sur un site forestier donné couvrant, la plupart du temps, 400 m<sup>2</sup>. Chaque relevé contient un inventaire des valeurs d'abondance/dominance de chaque espèce présente sur le site, selon l'échelle de [Braun-Blanquet \(1932\)](#), ainsi que de nombreuses variables écologiques pertinentes qui peuvent expliquer la distribution ou la diversité des plantes ou des unités de végétation (figure 1.1). Les variables écologiques proviennent soit de mesures directes sur le terrain (conditions topographiques, descriptions de sol, etc.), soit d'analyses effectuées en laboratoire (variables nutritionnelles chimiques et variables édaphiques physiques), soit de couches écologiques issues d'un système d'information géographique (variables climatiques). Avant d'être stockées dans la base de données, les variables sont standardisées, pour permettre l'intégration de données de différentes sources. Les protocoles d'acquisition des données sont aussi décrits dans divers champs d'EcoPlant et concernent principalement les méthodes d'échantillonnage des relevés, les caractéristiques des relevés floristiques et les méthodes d'analyse de sol ([Gégout et Jabiol 2001](#)).



**Figure 1.1.** Vue semi-réaliste d'un relevé particulier d'EcoPlant (VoS123), avec ses caractéristiques de végétation et de sol. Le dessin a été effectué par Cyril Galley.

Afin de pouvoir stocker de gros jeux de données répartis sur de vastes territoires et couvrant des gradients écologiques complets, des informations coûteuses en temps telles que des mesures de plantes individuelles, des comptages d'individus, ou des données nécessitant une instrumentation lourde sur le terrain, ne sont pas intégrées à EcoPlant. Des données qui peuvent varier sur une

base journalière ou hebdomadaire (comme par exemple des valeurs journalières de température ou de précipitations) ne sont pas non plus stockées, car elles ne sont valides qu'au niveau d'une étude particulière. En revanche, les variables écologiques qui varient temporellement à l'échelle de l'année ou plus (couvert des plantes, caractéristiques analytiques chimiques, etc.) sont stockées à chaque nouvelle observation du site, tandis que les variables stables avec le temps (altitude, exposition, topographie, etc.) peuvent n'être stockées qu'une seule fois par site.

La localisation géographique des sites, déterminée par leurs coordonnées géographiques, est calculée de manière systématique si nécessaire, puis intégrée à EcoPlant, pour pouvoir acquérir des données spatiales numérisées issues de croisements sous SIG. Aussi souvent que possible, une grande attention est portée à l'obtention de localisations précises des relevés par GPS, permettant ainsi des études diachroniques par des nouvelles mesures des variables pédologiques ou floristiques, sur le même site, mais à des périodes différentes.

### **1.3. Quelles variables pertinentes stocker dans EcoPlant ?**

#### ***1.3.1. Des caractéristiques complètes des sources des études***

EcoPlant est composée de jeux de données individuels, dont chacun représente une étude ou un projet particulier. Les références bibliographiques complètes et précises de ces études sont stockées dans EcoPlant, accompagnées d'informations sur les différents auteurs impliqués dans chaque étude, les organismes pour lesquels ils travaillent, ainsi que les institutions financières qui ont permis la réalisation des travaux. La stratégie d'échantillonnage utilisée pour collecter le jeu de données est aussi stockée à ce niveau. Il est ainsi possible de détecter les études pour lesquelles la stratégie d'échantillonnage a été gouvernée par la présence d'une espèce particulière, ou celles pour lesquelles aucune stratégie d'échantillonnage cohérente n'a été définie. Les auteurs des études, qui sont les fournisseurs des données, peuvent définir les conditions d'utilisation de leur données, et peuvent, par exemple, autoriser l'utilisation de leurs données uniquement avec leur permission explicite.

#### ***1.3.2. Des informations précises sur la localisation des relevés***

Les coordonnées géographiques, lorsqu'elles ne sont pas déterminées précisément dans le document source de l'étude, sont évaluées pour chaque relevé et intégrées dans EcoPlant. Selon l'information disponible pour un relevé particulier, un niveau de précision est affecté, avec un codage inspiré de celui de la base de données phytosociologiques SOPHY (Brisse et al. 1995). Chaque niveau de précision dans la localisation du relevé correspond à un type différent d'étude écologique ou biogéographique : (i) une précision de localisation de 100 km est suffisante pour certaines études biogéographiques dont l'objet est l'étude du comportement écologique des espèces végétales sur de vastes territoires (Diekmann et Lawesson 1999, Gégout et Krizova 2003, Coudun et Gégout 2005) ; (ii) une précision de localisation de 1 à 10 km est suffisante en contexte de plaine, pour des études où des variables climatiques directes peuvent être issues de stations météorologiques et d'analyses sous SIG (Austin et Meyers 1996) ; (iii) une précision de localisation de 50 m, possible lorsque les relevés sont localisés sur des cartes géographiques précises, est nécessaire lorsque des variables climatiques et méso-topographiques sont déterminées à partir de modèles numériques de terrain (MNT) pour comprendre par exemple la distribution des plantes en contexte de montagne (Allen et al. 1991, Franklin 1998, Guisan et al. 1998) ; (iv) enfin, une précision de localisation de quelques mètres, disponible lorsque les sites sont localisés avec un système de positionnement type GPS (global positioning system), permet aux auteurs de revenir sur les sites pour des examens complémentaires, et permet aussi le

monitoring des caractéristiques du sol ou l'installation permanente de relevés avec des mesures semi-quantitatives de l'abondance des espèces végétales.

### ***1.3.3. Des relevés floristiques dans une nomenclature cohérente et unifiée***

Un pré-requis nécessaire avant la constitution de toute base de données floristiques basée sur des études diverses, est l'accord préalable sur une nomenclature botanique. Les noms et les définitions des espèces utilisés par les auteurs peuvent varier et se référer à différentes flores, et ainsi conduire à l'utilisation de noms actuels différents pour une même espèce. Deux systèmes de nomenclature latine ont été adoptés pour EcoPlant : le "Code Informatisé de la la Flore de France" (CIFF, [Brisse et Kerguelen 1994](#)) pour les végétaux vasculaires, et la liste des bryophytes européens basées sur les travaux de [Corley et al. \(1981\)](#), ainsi que sur ceux de [Grolle \(1983\)](#). Les noms courants, les synonymes et leur correspondance à leur nom courant, sont présentés dans les deux systèmes taxonomiques, et la correspondance entre les systèmes CIFF et Flora Europaea ([Tutin et al. 2001](#)) a été effectuée. Il est ainsi possible de traduire tous les noms scientifiques latins dans les systèmes taxonomiques les plus communs. Les noms français et leur correspondance avec leur nom scientifique latin courant sont stockés dans EcoPlant, avec la référence aux flores les plus populaires de France ([Bonnier 1990](#), [Rameau et al. 1989, 1993](#)). Pour traiter les erreurs dans l'orthographe des noms d'espèces, des programmes semi-automatiques (macros sous Excel) ont été écrits, et un champ spécial dans EcoPlant est consacré à la collecte des noms d'espèces sans nom actuel équivalent, afin qu'une recherche complémentaire puisse être menée sur ces noms, par exemple en contactant directement les auteurs des observations, lorsque cela est possible.

En plus des problèmes taxonomiques rencontrés, la méthodologie adoptée sur le terrain pour réaliser les relevés floristiques est stockée dans EcoPlant. Les règles d'estimation de la surface des relevés, de description de la structure verticale des peuplements et de la notation du couvert des espèces, sont consignées dans des champs spécifiques. De nombreuses classifications des communautés, comme le classement phytosociologique ou le classement Corine, ainsi que le type de gestion du peuplement, sont intégrés à la base de données.

### ***1.3.4. Des données climatiques issues de croisements sous SIG***

Avoir accès à des données climatiques grâce à des mesures provenant de stations météorologiques est traditionnellement moins problématique que de sélectionner les variables les plus pertinentes qui expliquent au mieux la distribution de la végétation. Les données macroclimatiques telles que la température ou les précipitations peuvent potentiellement conduire à des centaines de variables (il suffit de considérer les valeurs minimales, maximales ou moyennes des valeurs mensuelles ou annuelles, ainsi que leurs combinaisons pour créer divers indices climatiques...). Dans EcoPlant, ces variables sont calculées par des manipulations sous SIG entre des placettes géo-référencées et des couches climatiques disponibles pour le territoire français (modèle Aurelhy, [Benichou et Le Breton 1987](#)), mais seules les données macroclimatiques les plus importantes (valeurs moyennes annuelles des températures et des précipitations, évapotranspiration et bilan hydrique) sont physiquement stockées dans la base de données, les autres variables et indices étant calculés de manière spécifique pour chaque étude envisagée. Des variables climatiques indirectes qui peuvent être mesurées sur le terrain, telles que la pente, l'exposition, la topographie ou l'altitude, sont également stockées dans la base de données. Afin de pouvoir estimer les conditions microclimatiques à travers la proportion de lumière atteignant la végétation de la strate herbacée, des estimations visuelles du pourcentage total de recouvrement de la végétation à 50 cm, 2 m ou au niveau de la strate arborée, sont stockées dans EcoPlant.

### ***1.3.5. Des données édaphiques issues de mesures sur le terrain et d'analyses en laboratoire***

Les données de description de sol sont stockées au niveau du profil et au niveau de l'horizon. Les informations au niveau de l'horizon, qui sont plus difficiles à traiter pour expliquer l'écologie des plantes, sont résumées par des variables synthétiques valables au niveau du profil complet. Les données descriptives de sol collectées sur le terrain pour chaque placette sont la géologie, le matériau parental, les types d'humus et de sol. Des informations complémentaires sont obtenues avec l'abondance et la profondeur d'apparition de l'effervescence, et avec des caractéristiques de l'hydromorphie qui sont directement pertinentes pour détecter un excès de calcium ou des conditions anaérobies.

Pour chaque placette, le premier horizon minéral au moins a été prélevé pour subir des analyses en laboratoire. Gégout et Jabiol (2001) ont suggéré une liste d'analyses physiques et chimiques pertinentes à effectuer pour des sols forestiers. Des analyses physiques telles que la détermination de la texture du sol (teneur en sables, limons et argile) permettent, quand elles sont couplées à l'information sur la teneur en cailloux des horizons et leur profondeur, de calculer la réserve utile en eau pour chaque profil (Jamagne et al. 1977). Les analyses chimiques de sols stockées dans EcoPlant donnent des informations sur (i) l'acidité (pH), (ii) la nutrition minérale (cations échangeables Ca, Mg, K, taux de saturation en bases échangeables), la nutrition azotée (C/N) et le phosphore disponible, ainsi que (iii) la toxicité pour les plantes (par exemple excès de  $Al^{3+}$  ou de  $CaCO_3$ ).

## **1.4. Chiffres clés d'EcoPlant**

### ***1.4.1. 200 champs floristiques et écologiques***

EcoPlant est une base de données relationnelle composée de 11 tables primaires interconnectées (figure 1.2). Ces tables permettent le stockage et la gestion de nombreuses informations floristiques (trois tables : Végétation, Relevés, Nomenclature), pédologiques (cinq tables : Profils, Horizons organiques, Horizons minéraux, Horizons analyse) et climatiques (une table : Placettes). Les informations relatives aux études sont stockées dans une table spécifique (Etudes) et la dernière table (Observations) est utilisée pour l'identification des relevés. Toutes les informations disponibles ont conduit à la définition de 200 champs pour chaque relevé (figure 1.2) : 15 champs concernent les références des études, 21 champs concernent l'identification et la localisation des relevés, 62 champs sont utilisés pour décrire les informations écologiques, climatiques, pédologiques et floristiques des relevés, 35 champs sont utilisés pour décrire chaque horizon du sol par observation sur le terrain, 15 champs décrivent les protocoles d'analyses des échantillons de sols en laboratoire, 43 champs permettent de stocker les variables issues des analyses en laboratoire, six champs décrivent les stratégies d'échantillonnage et les méthodes de description de la végétation, et enfin trois champs sont utilisés pour stocker le nom de chaque espèce présente dans chaque relevé, ainsi que la strate dans laquelle ladite espèce a été observée, et son coefficient d'abondance/dominance dans le relevé.

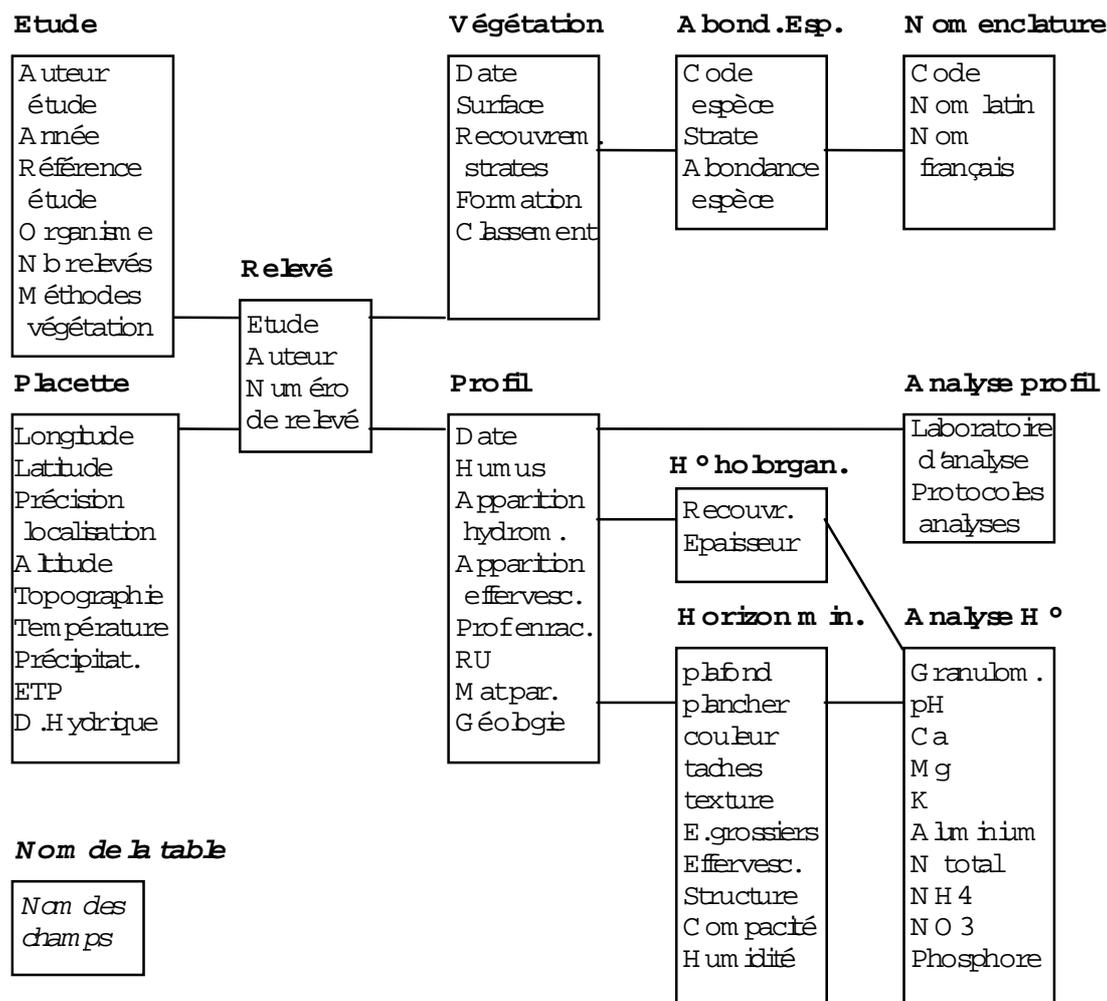


Figure 1.2. Structure simplifiée d'EcoPlant, avec 11 tables principales servant à stocker de l'information floristique et écologique complète au niveau d'une placette particulière. Seuls les champs écologiques les plus importants sont reportés. La figure est reprise de Gégout et al. (2002).

En pratique, l'interface informatique d'EcoPlant a été développée sous l'environnement informatique 4D (système de gestion de bases de données) et un moteur de recherches a été programmé pour permettre toutes les requêtes ou combinaisons de requêtes possibles au sujet de conditions floristiques ou écologiques particulières. Il est également facile d'importer de nouvelles données ou d'extraire des données sous la forme de fichiers texte qui peuvent ensuite être manipulés avec des logiciels statistiques ou des logiciels SIG, ou qui peuvent être incorporés dans d'autres systèmes de gestion. Enfin, un dictionnaire des données a été rédigé et diffusé.

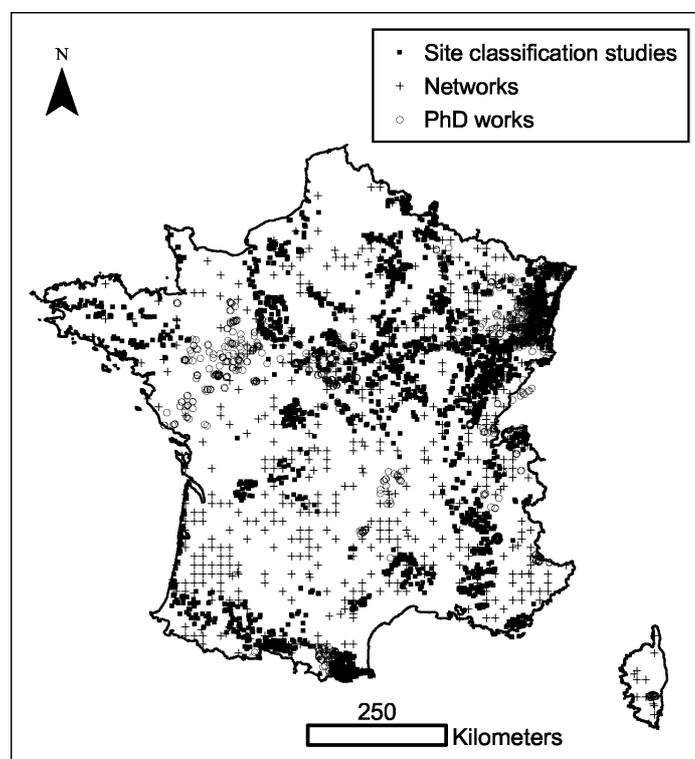
#### 1.4.2. 120 sources de relevés et 6 432 relevés phytoécologiques forestiers

Les 120 études qui sont actuellement stockées dans EcoPlant ont trois origines : 20 études sont le résultat de travaux de thèses de doctorat ou de maîtrise (données non publiées), deux études proviennent de réseaux de placettes forestières, soit au niveau national (RENECOFOR, Ulrich 1997) ou international (placettes du réseau européen de niveau I, Vanmechelen et al. 1997, Badeau 1998), et les 98 études restantes ont été réalisées dans le cadre d'un programme général de classification des stations forestières en France (tableau 1.1), qui a démarré dans les années 1980 et qui concerne une centaine de petites régions naturelles, couvrant deux tiers de la superficie de la France (figure 1.3). Pour chacune de ces régions naturelles, une classification numérique a été effectuée, sur la base de descriptions floristiques, édaphiques et environnementales de plusieurs centaines de relevés, afin de déterminer 20 à 50 groupes de relevés par région, appelés "types de

stations forestières". Pour chacun de ces groupes de stations, un relevé-type a été choisi, avec analyse de son sol en laboratoire. Enfin, une clé de détermination et une description précise de tous les types de stations et de tous les relevés-types sont fournies dans un rapport, le "catalogue de stations forestières" (Brêthes 1989). La totalité des relevés-types (descriptions de la végétation, du sol, du climat, etc.) des 98 catalogues de stations forestières disponibles en France ont été intégrés à EcoPlant.

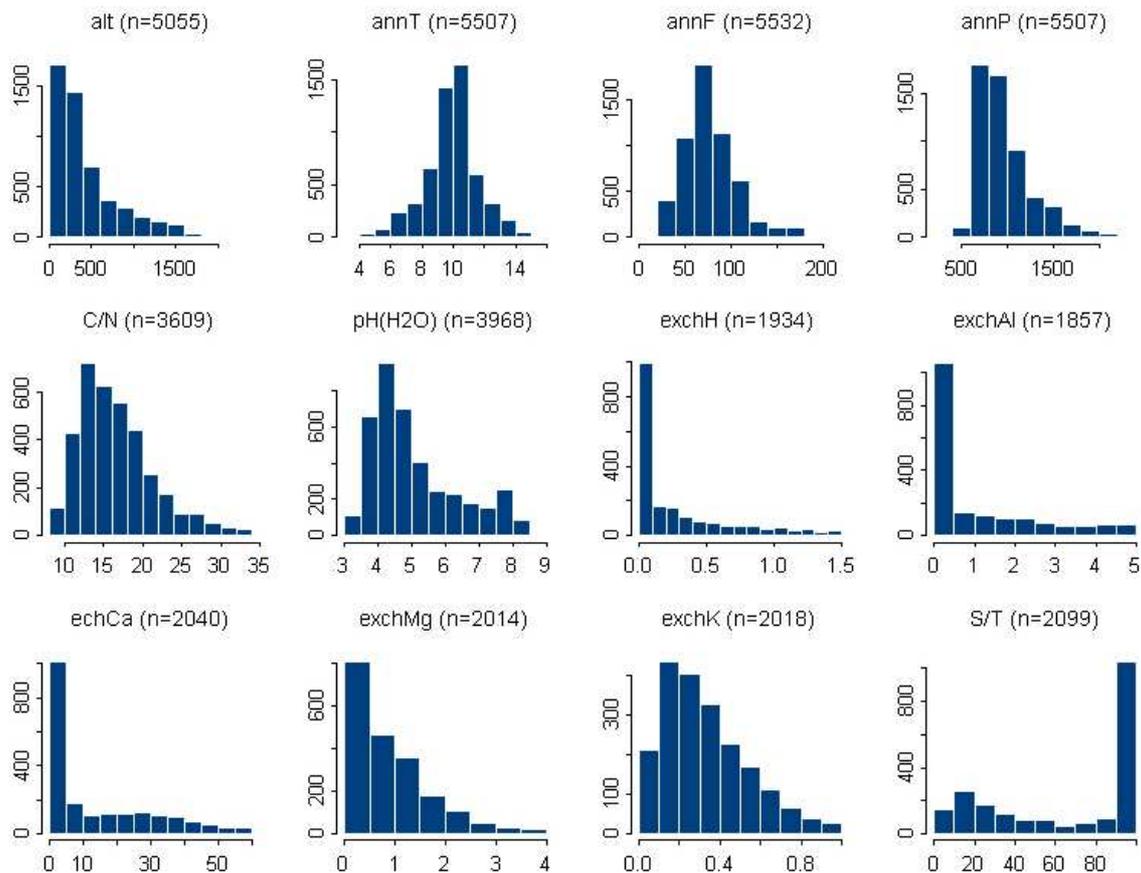
**Tableau 1.1.** Chiffres principaux d'EcoPlant, calculés pour chaque type de source de données. La plupart des relevés qui sont localisés précisément (< 10 m) ont été positionnés avec des instruments GPS GeoExplorer.

Sources de relevés	Etudes	Relevés floristiques	Descriptions de profils	Analyses chimiques	Précision de localisation < 10 m
Catalogues de stations	98	4 282	3 312	2 396	110
Thèses	20	1 507	736	1 497	562
Réseaux forestiers	2	643	642	642	0
<b>Total</b>	<b>120</b>	<b>6 432</b>	<b>4 690</b>	<b>4 535</b>	<b>672</b>



**Figure 1.3.** Localisation géographique des 6 432 relevés forestiers intégrés à EcoPlant, groupés par type de source de relevés.

Les 120 études stockées pour le moment dans EcoPlant contiennent 6 432 relevés phytoécologiques complets (tableau 1.1). Les relevés sont disséminés sur tout le territoire français (figure 1.3) et les variables écologiques, stockées avec les listes d'espèces associées, couvrent la gamme complète des conditions que l'on peut rencontrer dans les écosystèmes forestiers en France (figure 1.4). Des podzols jusqu'aux sols calcaires et des rankers jusqu'aux sols hydromorphes, la gamme des types et des caractéristiques de sols couvre une grande majorité des conditions de sols en Europe. La large gamme d'altitude (de 0 à plus de 2 000 m) et de conditions climatiques (de 4 à 18 °C de température moyenne annuelle et de 500 à 2 000 mm de précipitations moyennes annuelles) est également représentative des conditions climatiques dans une vaste zone européenne. Enfin, les relevés phytoécologiques intégrés dans EcoPlant concernent pour le moment des écosystèmes forestiers méditerranéens, océaniques et semi-continentaux, en conditions collinéennes ou montagnardes.



**Figure 1.4.** Conditions écologiques des placettes intégrées à EcoPlant, concernant quelques variables édaphiques et climatiques importantes.

alt : altitude (m) ; annT : température moyenne annuelle (°C) ; annF : nombre moyen annuel de jours de gel ; annP : précipitations moyennes annuelles (mm) ; C/N ; pH(H<sub>2</sub>O) : pH eau de l'horizon A du sol ; exchH : cations H<sup>+</sup> échangeables (CM+/kg), exchAl : aluminium échangeable (CM+/kg), exchCa : calcium échangeable (CM+/kg), exchMg : magnésium échangeable (CM+/kg), exchK : potassium échangeable (CM+/kg), S/T : taux de saturation au pH du sol, calculé par  $(\text{exchCa} + \text{exchMg} + \text{exchK}) / (\text{exchCa} + \text{exchMg} + \text{exchK} + \text{exchAl} + \text{exchH}) * 100$ . Les variables du sol qui sont présentées dans cette figure proviennent d'analyses en laboratoire du premier horizon organo-minéral du sol ; annT, annP et annF sont des variables climatiques calculées par croisements sous SIG, à partir du modèle Aurelhy. Les nombres entre parenthèses représentent le nombre de relevés pour lequel l'information écologique est disponible.

### 1.4.3. 2 154 taxons

En tout, 2 154 taxons différents ont été observés dans les 6 432 relevés, incluant ainsi très certainement la majorité des espèces végétales présentes dans les forêts de France. Néanmoins, parmi ces 2 154 taxons, seuls 364 sont présents dans plus de 1 % des relevés ; 13 espèces arborées (figure 1.5a) et 13 espèces arbustives (figure 1.5b) sont présentes dans plus de 10 % de tous les relevés ; 12 espèces herbacées et lianes (figure 1.5c) sont présentes dans plus de 15 % des relevés, et sept espèces muscinales (figure 1.5d) sont présentes dans plus de 10 % des relevés.

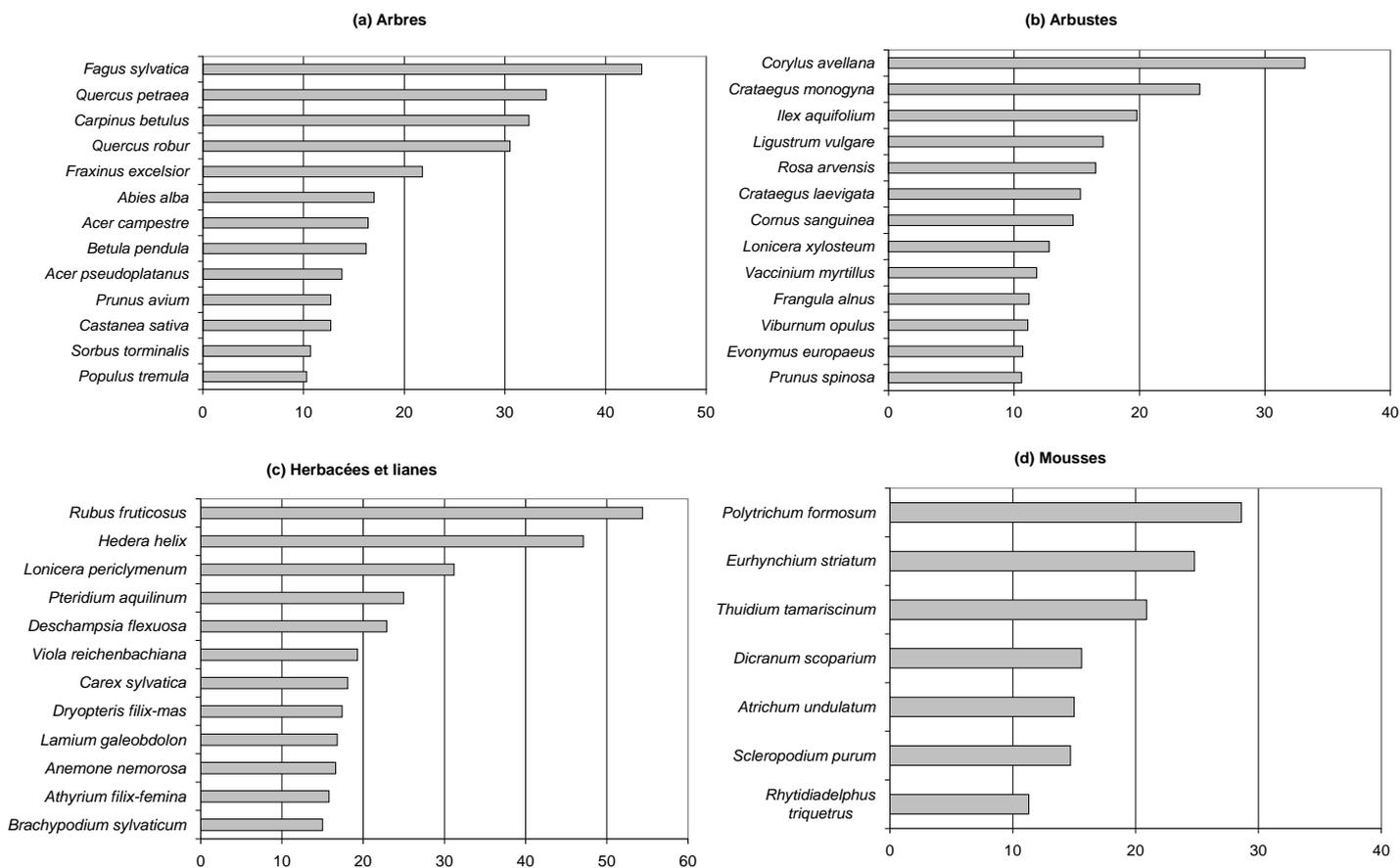


Figure 1.5. Fréquence (%) des espèces les plus fréquentes d'arbres (a), d'arbustes (b), d'herbes (c) et de mousses (d) intégrées dans EcoPlant.

Les espèces fréquentes présentent une grande importance, au vu de la distribution cumulée des fréquences d'espèces (figure 1.6) : le nombre d'occurrences des 10 espèces les plus fréquentes (*Hedera helix*, *Fagus sylvatica*, *Rubus fruticosus*, *Quercus petraea*, *Corylus avellana*, *Carpinus betulus*, *Lonicera periclymenum*, *Quercus robur*, *Polytrichum formosum* et *Pteridium aquilinum*) représente 14,7 % du nombre total d'occurrences. De la même manière, les 70 espèces les plus fréquentes représentent 50 % de toutes les occurrences et les 430 espèces les plus fréquentes représentent 90 % de toutes les occurrences.

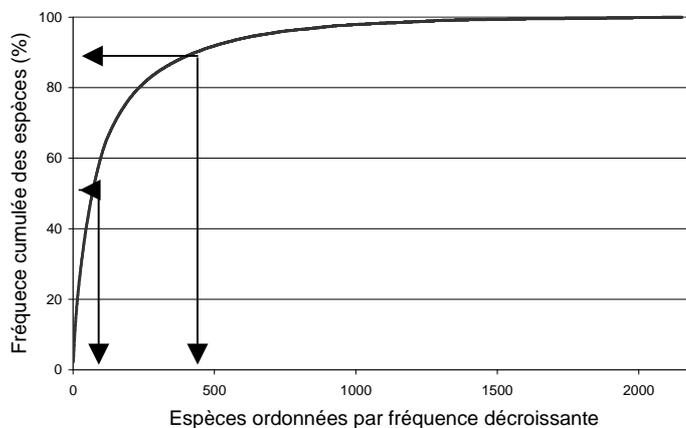


Figure 1.6. Fréquence cumulée des espèces végétales forestières en France, basée sur 6 432 relevés floristiques et 2 154 taxons, rangés par ordre de fréquence décroissante.

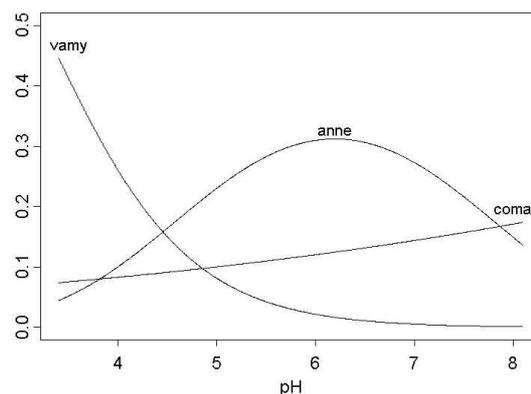
## 1.5. Conclusions et perspectives

### 1.5.1. Gestion des données et contrôle de la qualité

La plupart des études intégrées à EcoPlant étaient simplement disponibles dans une version papier (rapports, livres, etc.), et une attention particulière a été accordée à la gestion des données et au contrôle de leur qualité, à chaque étape de la constitution d'EcoPlant. Des règles cohérentes ont été définies, notamment pour les gammes de valeurs possibles des variables (par exemple 0 % < taux de saturation < 100 %). Une vérification systématique d'un relevé floristique et d'une analyse sur cinq a été effectuée pour chaque étude, pour éviter des erreurs systématiques et pour corriger ces erreurs. Enfin, un taux d'erreur entre les documents sources et les fichiers électroniques a été calculé, indépendamment des étapes précédentes, en vérifiant complètement un relevé sur 15, pour toutes les études. Le taux de données erronées, mesuré pour EcoPlant, est actuellement égal à 0,5 % et rend compte de la qualité des données.

### 1.5.2. Applications actuelles d'EcoPlant

EcoPlant pourra permettre l'étude des relations espèces-environnement et l'établissement de courbes de réponse écologique le long des gradients ayant le plus d'influence sur la distribution des espèces forestières végétales en France (voir par exemple la figure 1.7). Déjà EcoPlant a été utilisée pour comparer le comportement écologique de 46 espèces herbacées de forêt le long du gradient pH, dans deux régions du nord de la France (Coudun et Gégout 2005). La base de données a aussi été utilisée pour calculer des valeurs indicatrices écologiques, valables en France, pour 700 plantes forestières et six variables écologiques (Gégout et al. 2002). Par le biais d'une caractérisation écologique précise des cortèges phytosociologiques et des habitats naturels, EcoPlant pourra permettre une meilleure compréhension de l'organisation écologique des espèces et des communautés, ainsi que leur modélisation et leur spatialisation.



**Figure 1.7.** Courbes de réponse écologique (probabilités de présence) de *Vaccinium myrtillus* (vamy), *Anemone nemorosa* (anne) et *Cornus mas* (coma) le long du gradient pH, dans les forêts de France. 3 968 relevés d'EcoPlant ont été utilisés pour calculer ces courbes par des modèles de régression logistique.

### ***1.5.3. Perspectives***

EcoPlant a été conçue pour être compatible avec les principales bases de données floristiques et pédologiques de France. Un lien rapide peut être aussi effectué avec des bases de données de traits de vie ou d'attributs d'espèces. L'information contenue dans EcoPlant, référencée dans le temps et dans l'espace, avec des variables climatiques et des variables de ressources du sol, présente une importance fondamentale pour compléter les études actuelles de relations entre les plantes et le climat, dans un souci de modélisation géographique prédictive de la distribution des espèces.



## **2. Modélisation de la niche écologique des espèces végétales par la régression logistique et transposition des modèles dans l'espace géographique**

## Résumé

La régression logistique est présentée dans ce chapitre comme une technique de modélisation statistique efficace et robuste pour caractériser le comportement écologique des espèces forestières en France. Le principe mathématique de la technique, la méthode de sélection des variables pertinentes, et la définition d'un indice de qualité globale des modèles (D2) sont rappelés.

La représentation graphique des modèles est montrée sous forme de courbes et surfaces de réponse écologique et la réponse d'une espèce donnée à un facteur particulier peut être résumée par trois paramètres synthétiques dont la définition est donnée : l'optimum écologique, l'amplitude écologique et la probabilité maximale prédite le long du gradient. La transposition des modèles écologiques (probabilités de présence) à l'espace géographique est présentée, et un indice de qualité de prédiction spatiale du modèle est fourni (aire sous la courbe ROC).

La pertinence d'EcoPlant et des données qui y sont stockées justifie l'emploi de la régression logistique dans toute la thèse.

## Mots-clés

Régression logistique, courbe de réponse, surface de réponse, optimum écologique, amplitude écologique, carte de distribution potentielle, courbe ROC, EcoPlant.

## 2.1. Introduction

De nombreuses études en écologie végétale reposent sur la caractérisation quantitative des relations entre les espèces et leur environnement. L'approche statistique la plus commune en écologie, la régression linéaire par les moindres carrés, ne peut souvent pas être appliquée à des données de végétation, qui ne respectent pas les hypothèses de base de cette approche. En effet, pour une espèce donnée, la variable à expliquer est généralement sa présence/absence ou son couvert (abondance/dominance, pourcentage de recouvrement, etc.). La présence est notée 1 et l'absence est notée 0, avec des valeurs intermédiaires non permises, tout comme des valeurs inférieures à 0 ou supérieures à 1. Le pourcentage de recouvrement est également borné entre 0 et 100 % et les coefficients d'abondance/dominance de [Braun-Blanquet \(1932\)](#) sont exprimés dans une échelle ordinale (de + à 5). Des techniques statistiques inférentielles ont été développées pour tenir compte de la nature des variables qualitatives, parmi lesquelles on trouve les modèles linéaires généralisés, dont la régression logistique est un cas particulier ([McCullagh et Nelder 1997](#)).

La régression logistique permet de modéliser la relation entre une variable binaire à expliquer (0/1) et une ou plusieurs variables explicatrices le plus souvent quantitatives et étend donc les techniques de régression linéaire à des situations où la variable à expliquer ne suit pas obligatoirement une distribution normale. En pratique, les situations où la variable d'intérêt est catégorielle sont plutôt communes : succès/échec, amélioration/pas d'amélioration, présence/absence, etc.) et les applications de la régression logistique concernent de nombreux domaines tels que l'écologie ([Mladenoff et al. 1999](#)), la géologie, la sylviculture, la médecine (études épidémiologiques), les sciences sociales, etc. Dans toute la thèse, la régression logistique sera utilisée pour lier les données de présence/absence des espèces forestières avec les variables écologiques, afin d'obtenir une estimation de probabilité de présence des espèces en France pour chaque condition écologique.

Le développement récent des systèmes d'information géographique (SIG) a d'autre part permis de spatialiser les variables écologiques, le plus souvent climatiques, sous une forme numérique. Outre la manipulation aisée de couvertures thématiques concernant un même territoire, les SIG permettent la transformation des variables et l'intégration des équations logistiques liant la probabilité de présence des espèces aux variables environnementales spatialisées. Ces opérations rendent ainsi possible la spatialisation des modèles écologiques et l'obtention de cartes de distribution potentielle des espèces ([Guisan et al. 1998](#), [Guisan et Zimmermann 2000](#)).

L'objectif de ce chapitre est de présenter la technique de régression logistique dans le contexte de caractérisation quantitative du comportement écologique des plantes forestières en France.

## 2.2. Définition du modèle de régression logistique

La régression logistique suppose que la variable à expliquer  $Y$  soit binaire (1/0, présence/absence d'une espèce donnée). La technique ne modélise pas directement la variable  $Y$ , mais un rapport de probabilités associées aux valeurs de  $Y$ . On définit la proportion de la population pour les cas où  $Y = 1$  (présence de l'espèce) par  $p = p(Y = 1)$  et la proportion pour les cas où  $Y = 0$  (absence de l'espèce) par  $1 - p = p(Y = 0)$ . C'est le logarithme du rapport des probabilités de présence et d'absence ( $\ln(p/(1-p))$ , appelé transformation logit de  $p$ ) qui est modélisé. On suppose ensuite qu'un jeu de variables écologiques explicatrices  $X_1, X_2, \dots, X_p$ , est disponible. La régression consiste alors en la recherche d'une combinaison linéaire de ces variables écologiques pour ajuster le modèle (équation 2.1):

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_p \cdot X_p = \alpha + \sum_{j=1}^p \beta_j \cdot X_j \quad (\text{Equation 2.1})$$

où  $\alpha, \beta_1, \beta_2, \dots, \beta_p$ , sont les coefficients de la régression.

Il est simple ensuite, par une transformation exponentielle, de convertir le logarithme du rapport des probabilités (logit(p)) en probabilités de présence (équation 2.2a), ou en probabilités d'absence de l'espèce (équation 2.2b).

$$p = P(Y = 1) = \frac{e^{\alpha + \sum_{j=1}^p \beta_j \cdot X_j}}{1 + e^{\alpha + \sum_{j=1}^p \beta_j \cdot X_j}} = \frac{1}{1 + e^{-\alpha - \sum_{j=1}^p \beta_j \cdot X_j}} \quad (\text{Equation 2.2a})$$

$$1 - p = P(Y = 0) = \frac{1}{1 + e^{\alpha + \sum_{j=1}^p \beta_j \cdot X_j}} \quad (\text{Equation 2.2b})$$

Dans le contexte de régression logistique, les coefficients de régression  $\alpha$  et  $\beta_j$  sont estimés par le principe du maximum de vraisemblance et non sur le principe des moindres carrés. L'estimation du maximum de vraisemblance implique la définition de la fonction de vraisemblance L, comme le produit, pour toutes les observations, des probabilités de présence ou d'absence (équation 2.3).

$$L = \prod_{i=1}^n p_i = \prod_{i=1}^n \left[ \left( \frac{e^{\alpha + \sum_{j=1}^p \beta_j \cdot X_j}}{1 + e^{\alpha + \sum_{j=1}^p \beta_j \cdot X_j}} \right)^{Y_i} \cdot \left( \frac{1}{1 + e^{\alpha + \sum_{j=1}^p \beta_j \cdot X_j}} \right)^{1-Y_i} \right] \quad (\text{Equation 2.3})$$

où  $p_i$  est la probabilité du  $i^{\text{ème}}$  cas (pour lequel  $Y_i$  vaut soit 1, soit 0), et n est le nombre total d'observations. Le résultat de la fonction de vraisemblance L est souvent un très petit nombre et en pratique, on calcule son logarithme naturel (appelé LL) qui rend le résultat plus facile à manipuler. Puisque les probabilités sont inférieures à 1, les valeurs de logarithmes de vraisemblance sont toujours négatives et l'on considère deux fois l'opposé de ces valeurs de manière pratique, c'est la définition de la déviance (-2LL). L'estimation du maximum de vraisemblance revient à la maximisation de la fonction de vraisemblance, donc à la minimisation de la déviance (-2LL).

Pour un modèle donné, on peut tester le caractère significatif de chaque paramètre par la statistique de Wald, le rapport entre le coefficient et l'erreur résiduelle, qui suit une loi du t de Student. Il est donc possible de retenir les paramètres significatifs au seuil de 5 % ou de 1 %. Pour sélectionner les meilleures variables d'un jeu de données pour expliquer le comportement écologique d'une espèce donnée, on procède pas-à-pas, en intégrant les variables les plus significatives ou en supprimant les variables les moins significatives. Trois critères communs sont utilisés pour comparer des modèles : le critère de déviance (-2LL, équation 2.3) d'une part, et les critères d'Akaike (AIC, équation 2.4a, Akaike 1973) et de Schwartz (BIC, équation 2.4b, Schwartz 1978) d'autre part, qui tendent à pénaliser les modèles qui intégreraient trop de paramètres :

$$AIC = -2LL + 2 \cdot (1 + n_{\text{var}}) \quad (\text{Equation 2.4a})$$

$$BIC = -2LL + \ln(n) \cdot (1 + n_{\text{var}}) \quad (\text{Equation 2.4b})$$

où  $-2LL$  est la déviance du modèle,  $n_{\text{var}}$  est le nombre de variables explicatrices, le 1 correspond au coefficient  $\alpha$ , et  $n$  est le nombre total d'observations (nombre de relevés). Le critère de Schwartz pénalise fortement les modèles qui contiennent un plus grand nombre d'observations ou de variables (Coudun 2001).

Pour caractériser la qualité globale d'un modèle, il n'existe pas d'analogie directe au coefficient d'ajustement des moindres carrés ( $R^2$ ), car cette statistique quantifie un pourcentage de variance expliquée, et la variance d'une variable binaire dépend de la fréquence de distribution de cette variable. Néanmoins, une mesure de la force de l'association entre la variable binaire et les variables écologiques, est communément utilisée. Le D2, varie également entre 0 et 1, et indique la capacité des variables explicatrices à prédire la variable expliquée (équation 2.5).

$$D2 = \frac{dev_{\text{null}} - dev}{dev_{\text{null}}} \quad (\text{Equation 2.5})$$

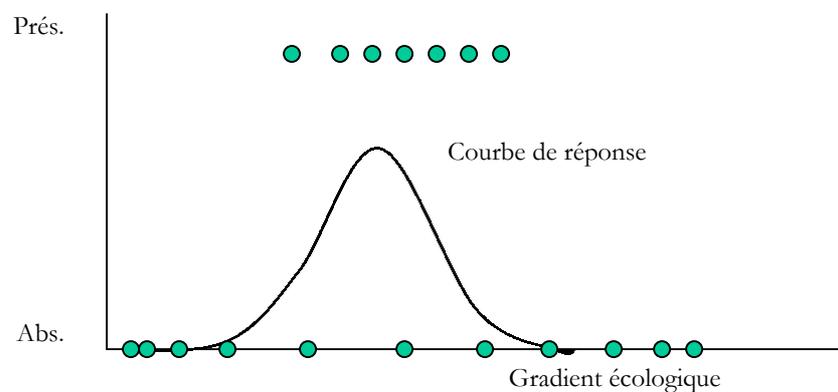
où  $dev$  est la déviance du modèle et  $dev_{\text{null}}$  la déviance nulle ( $-2LL$  pour le modèle qui n'inclut qu'une constante sans variable explicatrice). Ainsi, D2 correspond à la proportion de déviance expliquée sur la déviance totale.

De manière pratique, tous les calculs de régression logistique présentés dans la thèse ont été générés par le logiciel R/S-Plus (MathSoft 1999).

## 2.3. Interprétation des modèles dans l'espace écologique : les courbes et surfaces de réponse

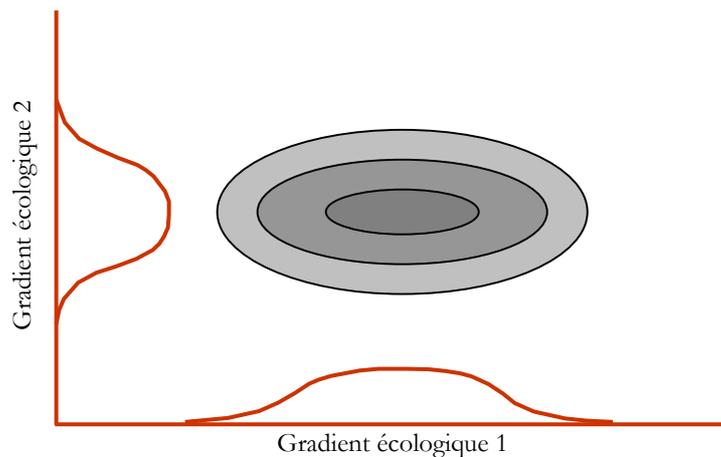
### 2.3.1. Courbes et surfaces de réponse écologique

Lorsque l'on effectue une régression logistique entre la présence/absence d'une espèce et une seule variable écologique (un gradient écologique), le résultat fournit des valeurs de probabilités estimées de présence de l'espèce pour chaque valeur de la variable. Représenter cette probabilité de présence d'une espèce le long d'un gradient écologique revient à tracer sa courbe de réponse écologique (figure 2.8).



**Figure 2.8.** Exemple schématique d'une courbe de réponse (probabilité de présence d'une espèce donnée) le long d'un gradient écologique. La présence/absence réelle de l'espèce est illustrée par les points grisés.

De la même manière, une surface de réponse est la représentation de la réponse écologique d'une espèce le long de deux gradients considérés simultanément (figure 2.9).

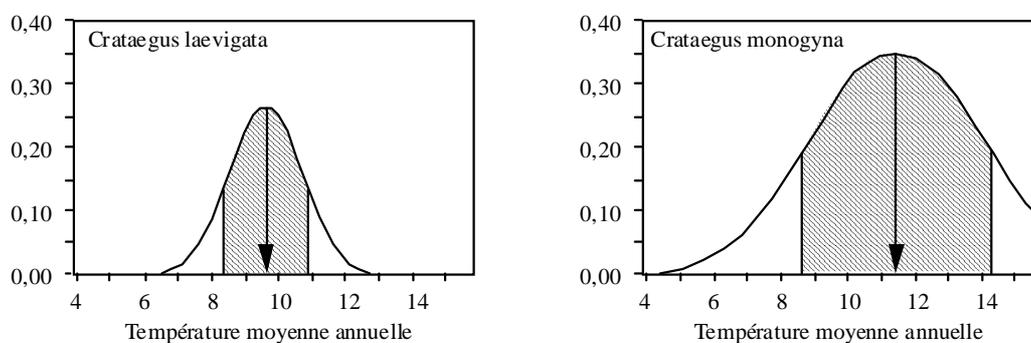


**Figure 2.9.** Exemple schématique d'une surface de réponse le long de deux gradients écologiques. Les deux courbes de réponse par rapport à chacune des variables sont également dessinées.

### 2.3.2. Optimum et amplitude écologiques

La courbe de réponse écologique d'une espèce par rapport à un gradient fournit la probabilité de présence de l'espèce pour chaque valeur du gradient. Divers résumés numériques sont traditionnellement extraits des courbes de réponse, pour synthétiser l'information portée par la courbe complète : on définit alors (i) l'optimum écologique, la valeur de la variable pour laquelle la probabilité est maximale, (ii) l'amplitude écologique, caractérisant la gamme plus ou moins étendue des milieux dans lesquels l'espèce peut survivre, et (iii) la probabilité maximale de la courbe, caractérisant la fréquence maximale de l'espèce le long du (ou des) gradient(s).

Si l'optimum et la probabilité maximale de la courbe ont des définitions claires et sans ambiguïté, l'amplitude quant à elle peut avoir plusieurs définitions : [ter Braak et Looman \(1986\)](#) ont donné une définition de la tolérance calculée à partir des coefficients de la régression logistique, mais qui ne s'applique qu'à des courbes unimodales (en forme de cloche) ; [Gégout et Pierrat \(1998\)](#) ont donné la définition de l'amplitude comme un certain pourcentage de l'aire sous la courbe de réponse pour les probabilités maximales ; enfin [Heegaard \(2002a\)](#) a donné une définition de l'amplitude basée sur une distance entre les deux points de la courbe de réponse situés à une probabilité de présence équivalente à une proportion fixée de la probabilité maximale.



**Figure 2.10.** Courbe de réponse unimodale (en cloche), optimum et amplitude écologiques de *Crataegus laevigata* et *Crataegus monogyna* vis-à-vis de la température moyenne annuelle. La zone correspondant à l'amplitude écologique est hachurée. L'optimum écologique (valeur indicatrice) de l'espèce, est indiqué par une flèche (source : [Gégout et al. 2002](#)).

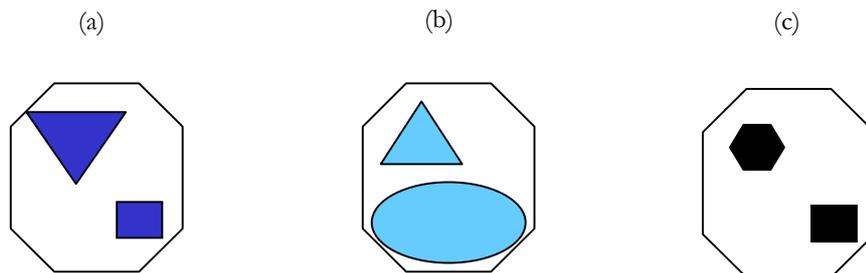
Parmi ces trois définitions, c'est celle de [Gégout et Pierrat \(1998\)](#) que nous avons appliquée dans toute la thèse, puisqu'elle s'applique à toutes les formes de courbes de réponse. L'indice correspond à un domaine de la variable étudiée, éventuellement constitué de parties disjointes, où en tous points, la fréquence de l'espèce est supérieure à celle des points du domaine non retenu (figure 2.10). L'avantage est que le domaine retenu correspond à une part de distribution égale pour toutes les espèces ([Gégout et al. 2002](#)), fixée à 80 % dans le cadre de cette thèse.

## 2.4. Interprétation des modèles dans l'espace géographique : les cartes de distribution potentielle

La méthodologie présentée ci-dessous, pour comparer les probabilités de présence prédites par les modèles de régression logistique et les présences/absences réelles des espèces, est tout à fait généralisable aux autres techniques de modélisation des relations espèces/environnement dont le résultat fournit des valeurs de probabilité de présence (modèles additifs généralisés, [Yee et Mitchell 1991](#), [Bio et al. 1998](#), arbres de classification, [Vayssières et al. 2000](#), etc.).

### 2.4.1. Spatialisation des modèles

A chaque condition écologique (combinaison de valeurs des variables écologiques), on peut associer une probabilité de présence. En remplaçant les variables écologiques du modèles par les cartes correspondantes, il est aisé de prédire spatialement la probabilité de présence d'une espèce à l'échelle d'une région ou d'un pays (figure 2.11).



**Figure 2.11.** Carte schématique des habitats favorables pour une espèce donnée pour (a) une variable 1 uniquement, (b) une variable 2 uniquement, et (c) les variables 1 et 2 considérées simultanément.

### 2.4.2. Evaluation des cartes de distribution potentielle

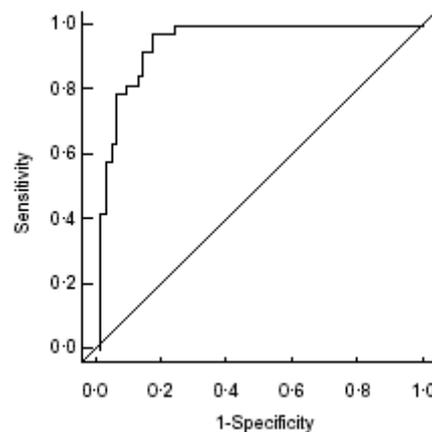
L'évaluation de la carte de distribution potentielle d'une espèce consiste en la comparaison des probabilités de présence prédites (variable continue dont les valeurs sont comprises entre 0 et 1) avec les observations réelles de présence/absence (variable binaire 0/1). Pour transformer le vecteur de probabilités de présence en un vecteur de données binaires de présence/absence prédite, on peut choisir un seuil arbitraire (souvent 0,5) au-dessus duquel on affecte la valeur 1 (présence prédite) et au-dessous duquel on affecte 0 (absence prédite). Ainsi, on peut comparer deux vecteurs de données binaires et définir une matrice de confusion (tableau 2.2).

**Tableau 2.2.** Définition de la matrice de confusion, table de contingence entre un vecteur de présence/absence observée et un vecteur de présence/absence prédite.

	Présence prédite	Absence prédite
Présence observée	a	c
Absence observée	b	d

L'établissement d'une telle matrice de confusion permet de calculer différentes statistiques dont les principales sont : (i) le succès global de la prédiction ( $S = (a+d)/n$ , où  $n$  est le nombre total de relevés,  $n = a+b+c+d$ ) ; (ii) la sensibilité, ou pourcentage de bonne prédiction de la présence observée ( $S_n = a/(a+c)$ ) ; (iii) la spécificité, ou pourcentage de bonne prédiction de l'absence observée ( $S_p = d/(b+d)$ ).

Cependant, cette approche est dépendante du seuil de probabilité de présence choisi (ici 0,5) pour l'affectation des présences (1) et absences (0) prédites. Une technique récente, empruntée à la télécommunication, permet de s'affranchir de l'influence du seuil de probabilité en testant tous les seuils possibles (de 0 à la probabilité maximale prédite), et de tracer une courbe ROC (receiver-operating characteristic), qui représente  $S_n$  en fonction de  $1 - S_p$  (voir [Fielding et Bell 1997](#) et la figure 2.12).



**Figure 2.12.** Exemple de courbe ROC (source : [Manel et al. 2001](#)).

Un indice de qualité de prédiction du modèle considéré est l'aire sous la courbe ROC (AUC = area under the curve), et une échelle empirique a été proposée pour qualifier le modèle :  $0,9 < AUC < 1,0$  : excellent modèle ;  $0,8 < AUC < 0,9$  : bon modèle ;  $0,7 < AUC < 0,8$  : modèle correct ;  $0,6 < AUC < 0,7$  : modèle médiocre ;  $0,5 < AUC < 0,6$  : mauvais modèle.

Le seuil de probabilité sélectionné, permettant d'établir une matrice de confusion optimale en terme de succès de prédiction à la fois de la présence et de l'absence observée, correspond au point le plus éloigné de la diagonale de la figure 2.12, puisque c'est le point où le couple d'indices ( $S_n$ ,  $S_p$ ) est le meilleur.

## 2.5. Discussion

La régression logistique est une technique flexible, paramétrique, facilement programmable avec des logiciels de statistiques. Tous les calculs de la thèse ont été réalisés avec le logiciel S-Plus (MathSoft 1999). Il existe de nombreuses autres techniques pour lier la présence/absence d'une espèce aux conditions écologiques (Segurado et Araújo 2004), mais la régression logistique s'est montrée efficace dans de nombreuses études.

C'est cette technique de modélisation que nous avons utilisée tout au long de la thèse pour établir le lien entre les espèces et les variables écologiques. Celles-ci ont été classées par Austin (Austin et al. 1984, Austin et Smith 1989) en trois grandes catégories : (i) les gradients de ressource, représentant la matière et l'énergie consommés par les plantes (par exemple nutriments, eau, lumière), (ii) les gradients de conditions (gradients directs), ayant une importance physiologique, mais qui ne sont pas consommés (par exemple température, pH), (iii) les gradients indirects, qui n'ont pas de pertinence physiologique directe sur les organismes (pente, aspect, altitude, position topographique, type d'habitat, géologie, etc.). Pour être compréhensibles et fonctionnels, les modèles de relation entre plantes et environnement doivent privilégier les gradients directs et les gradients de ressources.

La base de données EcoPlant présentée dans le chapitre précédent stocke les variables de ressources et de conditions les plus facilement accessibles. Elle peut donc se révéler être un outil précieux et pertinent pour caractériser la réponse des espèces forestières à des facteurs du climat et du sol.



**3. L'établissement de courbes de réponse écologique avec la régression logistique est sensible à l'intensité de l'échantillonnage et aux caractéristiques des courbes**

*Article associé*

**Coudun, Ch. et J.-C. Gégout (en révision mineure)** The derivation of species response curves with logistic regression models is sensitive to sampling intensity and curve characteristics. *Ecological Modelling*.

## Résumé

Nous avons caractérisé de manière quantitative la sensibilité des courbes de réponse écologique des espèces aux caractéristiques d'échantillonnage (nombre de relevés, nombre d'occurrences et prévalence des espèces), le long d'un gradient de pH artificiel. Nous avons défini 54 courbes théoriques de réponse unimodale, issues de la combinaison de six valeurs d'optimum ( $opt = 3$ ,  $opt = 4, \dots, opt = 8$ ), trois valeurs de tolérance ( $tol = 0,5$ ,  $tol = 1,0$  et  $tol = 1,5$ , *sensu* ter Braak et Looman 1986), et trois valeurs pour la probabilité de présence maximale ( $p_{max} = 0,05$ ,  $p_{max} = 0,20$  et  $p_{max} = 0,50$ ). Pour chacune de ces courbes de réponse théoriques, nous avons construit des jeux de données binaires artificiels (présence/absence), afin de tester l'influence de l'occurrence des espèces, de leur prévalence, ainsi que du nombre de relevés disponibles. A l'aide de données réelles extraites d'EcoPlant, base de données phytoécologiques pour les forêts de France (Gégout et al. 2005), nous avons comparé la réponse écologique de 50 plantes forestières au pH du sol, en se basant tout d'abord sur un petit jeu de données (100 relevés échantillonnés au hasard), puis en se basant sur toutes les données disponibles dans EcoPlant (3810 relevés).

Les résultats obtenus avec les données artificielles ont montré que l'optimum, l'amplitude, ou la probabilité maximale de la courbe de réponse, sont difficiles à déterminer de manière fiable avec la régression logistique, lorsqu'une espèce est trop peu fréquente, ou lorsque son optimum théorique se trouve près d'une extrémité du gradient. Ces résultats théoriques ont été illustrés avec les données réelles extraites d'EcoPlant. Nous proposons une valeur minimale de 50 à 100 occurrences de l'espèce, pour que sa courbe de réponse écologique soit calculée de manière fiable avec la régression logistique.

## Mots-clés

Données artificielles, amplitude écologique, optimum écologique, base de données EcoPlant, plantes forestières, régression logistique, pH, courbe de réponse, échantillonnage.

### 3.1. Introduction

La possibilité de lier la présence/absence des espèces à des facteurs écologiques par l'établissement de courbes de réponse a suscité l'intérêt des écologues ces dernières décennies (ter Braak et Looman 1986, Odland et al. 1995, ter Braak 1996, Austin 2002). Les techniques statistiques les plus utilisées actuellement sont les modèles linéaires généralisés (GLMs, McCullagh et Nelder 1997) et les modèles additifs généralisés (GAMs, Yee et Mitchell 1991, Hastie et Tibshirani 1997) et ces techniques ont été très documentées, à la fois sur des aspects théoriques et empiriques (Guisan et al. 2002, Lehmann et al. 2002a, Scott et al. 2002). La régression logistique est un cas particulier de modèle linéaire généralisé, et l'une des plus anciennes techniques et des plus utilisées pour lier la présence/absence des espèces à des gradients écologiques. Elle permet de caractériser quantitativement les relations espèces/environnement de manière simple, flexible et paramétrique, ce qui suppose des courbes de réponse écologique en forme de cloche (ter Braak et Looman 1986, ter Braak 1996). Lorsque l'objectif consiste à résumer numériquement l'information portée par les courbes de réponse écologique, pour obtenir par exemple l'optimum et l'amplitude écologiques des espèces, la régression logistique s'est révélée être une technique robuste et puissante (Hill et al. 1999, 2000, Roy et al. 2000, Gégout et Krizova 2003, Coudun et Gégout 2005). Dans une synthèse récente, Diekmann (2003) souligne que la régression logistique est tout à fait appropriée pour calculer des valeurs indicatrices pour les espèces végétales.

La décision de sélectionner ou de rejeter une espèce d'un jeu de données, dans l'optique d'un calcul d'une courbe de réponse écologique, est une étape importante pour les écologues qui souhaitent caractériser le comportement écologique de nombreuses espèces dans une région particulière, mais leur décision est rarement expliquée ou justifiée (McKenney et al. 2002). Dans la plupart des études, les espèces sélectionnées sont celles qui sont présentes dans le jeu de données dans un nombre ou une proportion de relevés supérieur à un seuil arbitraire (tableau 3.3). Ce seuil minimal d'occurrence est très variable selon les études, puisqu'on rencontre des valeurs comprises par exemple entre 5 (Guisan et al. 1999, Austin et al. 2000, Araújo et al. 2002, Cawsey et al. 2002, Gégout et al. 2003, Segurado et Araújo 2004) et 100 occurrences (Dirnböck et al. 2003, Dirnböck et Dullinger 2004), tandis que la prévalence minimale pour sélectionner une espèce est encore plus variable, puisqu'elle dépend de la taille du jeu de données (Manel et al. 2001, tableau 3.3). Lorsque l'objectif de l'étude est d'étudier le comportement écologique de nombreuses espèces dans une région (voir par exemple Lawesson et Oksanen 2002, Coudun et Gégout 2005), le seuil d'occurrence minimal pour sélectionner une espèce est très important puisque, dans la nature, on rencontre peu d'espèces fréquentes et beaucoup d'espèces peu fréquentes (Karl et al. 2002). Le choix d'un seuil plus bas peut donc augmenter le nombre d'espèces potentiellement étudiées, mais les conséquences sur les courbes de réponse restent très mal connues.

**Tableau 3.3.** Liste de 21 études de modélisation de la distribution des espèces, basées sur les techniques GLM/GAM. Des renseignements sont donnés sur le lieu de l'étude, la nature des espèces modélisées, ainsi que sur les caractéristiques de l'échantillonnage (nombre de relevés, nombre d'espèces, occurrence minimale des espèces et prévalence minimale des espèces). Le critère qui a servi à sélectionner les espèces est indiqué en gras pour chaque étude.

Référence	Zone d'étude	Espèces	Nombre de relevés	Nombre d'espèces	Occurrence minimale	Prévalence minimale
<a href="#">Araújo et al. (2002)</a>	Grande-Bretagne	Oiseaux	120 000	78	<b>5</b>	0,0 %
<a href="#">Araújo et al. (2004)</a>	Europe	Plantes	2 434	1 200	<b>25</b>	1,0 %
<a href="#">Austin (1998)</a>	Australie	Arbres	9 537	88	<b>50</b>	0,5 %
<a href="#">Austin et al. (2000)</a>	Australie	Arbres et arbustes	2 530	135	<b>5</b>	0,2 %
<a href="#">Bakkenes et al. (2002)</a>	Europe	Plantes	4 419	1 397	<b>20</b>	0,5 %
<a href="#">Bio et al. (1998)</a>	Pays-Bas	Plantes	2 090	156	<b>22</b>	1,1 %
<a href="#">Bio et al. (2002)</a>	Belgique	Plantes	2 587	18	<b>20</b>	0,8 %
<a href="#">Bragazza et Gerdol (1996)</a>	Italie	Plantes	251	21	<b>25</b>	<b>10,0 %</b>
<a href="#">Brotons et al. (2004)</a>	Espagne	Oiseaux	1 550	30	<b>15</b>	1,0 %
<a href="#">Cawsey et al. (2002)</a>	Australie	Arbres et arbustes	2 307	147	<b>5</b>	0,2 %
<a href="#">Coudun et Gégout (2005)</a>	France	Plantes	1 200	46	<b>50</b>	4,2 %
<a href="#">Dirnböck et al. (2003)</a>	Autriche	Plantes	959	85	<b>100</b>	10,4 %
<a href="#">Dirnböck et Dullinger (2004)</a>	Autriche	Plantes	1 016	71	<b>100</b>	9,8 %
<a href="#">Gégout et al. (2003)</a>	France	Plantes	306	122	<b>5</b>	1,6 %
<a href="#">Guisan et al. (1999)</a>	Etats-Unis	Arbres et arbustes	144	23	<b>4</b>	2,8 %
<a href="#">Guisan et Theurillat (2000)</a>	Suisse	Plantes	205	62	<b>15</b>	7,3 %
<a href="#">Heegaard et al. (2001)</a>	Irlande du Nord	Plantes aquatiques	574	32	<b>20</b>	3,5 %
<a href="#">Lehmann et al. (2002b)</a>	Nouvelle-Zélande	Fougères	19 875	43	<b>202</b>	<b>1,0 %</b>
<a href="#">McPherson et al. (2004)</a>	Afrique du Sud	Oiseaux	4 275	32	<b>26</b>	<b>0,6 %</b>
<a href="#">Segurado et Araújo (2004)</a>	Portugal	Amphibiens/Reptiles	993	44	<b>4</b>	<b>0,4 %</b>
<a href="#">Zaniewski et al. (2002)</a>	Nouvelle-Zélande	Fougères	19 875	43	<b>66</b>	0,3 %

La position de l'optimum de l'espèce le long du gradient écologique semble d'autre part avoir une influence sur le calcul des courbes de réponse, et certains artéfacts ont été mis en évidence, tels que l'effet de l'échantillonnage sur la distribution des espèces le long des gradients ([Mohler 1983](#)). Cela peut rendre difficile le calcul des optima proches des extrémités des gradients ([Rydgren et al. 2003](#)). [Austin et al. \(1990, 1994\)](#) ont également montré que les espèces avaient tendance à présenter plus de courbes asymétriques lorsque leur optimum se trouve proche d'une extrémité du gradient.

Grâce à la création d'un jeu de données, notre objectif principal dans ce chapitre était d'évaluer un nombre minimal d'occurrences nécessaire pour sélectionner une espèce et calculer une courbe de réponse fiable avec la régression logistique. Nous voulions également illustrer certains artéfacts techniques liés à la régression logistique, qui peuvent expliquer en partie la distribution observées des optima, ainsi que la proportion d'espèces apparemment indifférentes, en considérant des jeux de données réels issus de la base de données EcoPlant.

## 3.2. Matériels et méthodes

### 3.2.1. Régression logistique et résumés numériques des courbes de réponse

Une courbe de réponse représente la probabilité de présence  $p(x)$  d'une espèce le long d'un gradient écologique  $x$ , et sa forme conceptuelle est unimodale (en forme de cloche, [ter Braak 1996](#)), ce qui signifie, dans une approche par la régression logistique, que  $p(x)$  soit liée, sous une forme transformée, à une fonction quadratique de  $x$  (équation 3.6, de [ter Braak et Looman 1986](#)):

$$\log\left(\frac{p(x)}{1-p(x)}\right) = b_0 + b_1 \cdot x + b_2 \cdot x^2 = a - \frac{(x - opt)^2}{2 \cdot tol^2} \quad \text{(Equation 3.6)}$$

Dans l'équation 3.6,  $b_0$ ,  $b_1$  et  $b_2$  sont les paramètres de régression issus d'un calcul exécuté avec un logiciel de statistique tel que S-Plus ([MathSoft 1999](#)), et  $opt$ ,  $tol$  et  $a$  sont respectivement l'optimum, la tolérance et un paramètre lié à  $p_{max}$ , la valeur maximale de  $p(x)$ . Les trois paramètres  $opt$ ,  $tol$ , et  $p_{max}$  peuvent être facilement liés aux coefficients de régression (équation 3.7, de [ter Braak et Looman 1986](#)).

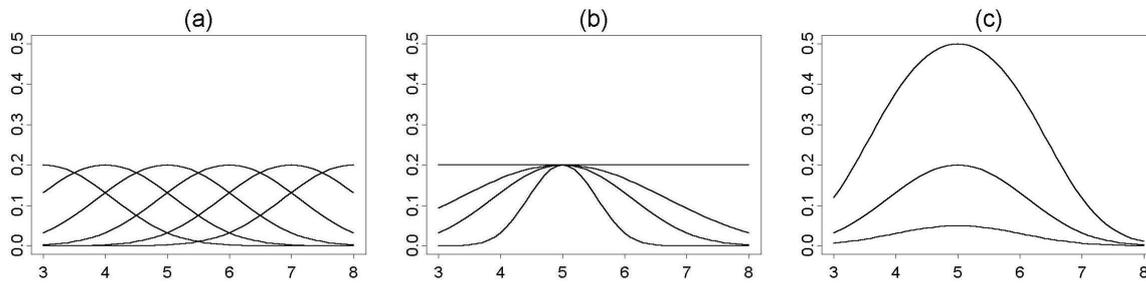
$$opt = \frac{-b_1}{2 \cdot b_2}, \quad tol = \frac{1}{\sqrt{-2 \cdot b_2}}, \quad p_{max} = \frac{1}{1 + \exp\left(\frac{b_1^2}{4 \cdot b_2} - b_0\right)} \quad \text{(Equation 3.7)}$$

Lorsque la courbe de réponse le long d'un gradient n'est pas unimodale de manière significative, ce qui se détermine par un test de déviance au seuil par exemple de 0,05 pour  $p$  ([McCullagh et Nelder 1997](#)), l'équation 3.6 peut être simplifiée avec  $b_2$  égal à zéro, et la courbe de réponse est alors sigmoïde, croissante ou décroissante. De la même manière, lorsque la courbe de réponse n'est pas non plus sigmoïde de manière significative, l'équation 3.6 est simplifiée avec  $b_1$  et  $b_2$  égaux à zéro et la courbe de réponse est plate (pas de réaction apparemment significative de l'espèce vis-à-vis du gradient écologique). Le calcul des résumés numériques  $opt$ ,  $tol$ , et  $p_{max}$  à partir de l'équation 3.7, n'est valide que pour des courbes de réponse unimodales en forme de cloche ( $b_2$  est significativement différent de zéro), et permet la comparaison du comportement écologique d'espèces différentes le long des gradients ([Odland et al. 1995](#)), ou la comparaison du comportement écologique des mêmes espèces, mais dans des régions différentes ([Gégout et Krizova 2003](#), [Coudun et Gégout 2005](#)).

### 3.2.2. Courbes de réponse théoriques

Nous avons utilisé un gradient réel de pH pour explorer la sensibilité de la courbe de réponse des espèces à l'intensité de l'échantillonnage et aux caractéristiques de la courbe. Dans les forêts françaises, les valeurs de pH rencontrées se situent entre 3 et 8, pour la plupart des sites intégrés à la base de données forestières EcoPlant ([Gégout 2001](#), [Gégout et al. 2005](#)). Nous avons défini 54 courbes de réponse théoriques, représentatives des courbes de réponse de plantes réelles le long du gradient pH en France ([Coudun et Gégout 2005](#)), comme combinaisons de six valeurs pour l'optimum ( $opt = 3, opt = 4, \dots, opt = 8$ ), trois valeurs pour la tolérance ( $tol = 0,5, tol = 1,0, tol = 1,5$ ), et trois valeurs pour la probabilité maximale de la courbe ( $p_{max} = 0,05, p_{max} = 0,20, p_{max} = 0,50$ , voir l'équation 3.7 et la figure 3.13). Ces 54 courbes de réponse peuvent être

aisément interprétées comme la représentation du comportement écologique de 54 plantes virtuelles le long du gradient pH (figure 3.13), avec des espèces acidiphiles ( $opt = 3$ ) à calcariques ( $opt = 8$ ), des espèces à faible ( $tol = 0,5$ ) à large ( $tol = 1,5$ ) amplitude, et des espèces peu fréquentes ( $pmax = 0,05$ ) à fréquentes ( $pmax = 0,50$ ).



**Figure 3.13.** Probabilité de présence théorique des espèces simulées le long du gradient de pH. Les 54 courbes de réponse simulées sont issues de combinaisons de (a) six valeurs pour l'optimum ( $opt = 3, 4, 5, 6, 7, \text{ ou } 8$ ), (b) trois valeurs pour la tolérance ( $tol = 0,5, 1,0 \text{ ou } 1,5$ ), et (c) trois valeurs pour la probabilité maximale de la courbe ( $pmax = 0,05, 0,20 \text{ ou } 0,50$ ). La figure 3.13b montre également une courbe de réponse plate (pas de réponse attendue pour l'espèce le long du gradient).

De plus, nous avons défini trois courbes de réponse théoriques plates ( $pmax = 0,05$ ,  $pmax = 0,20$ , et  $pmax = 0,50$ ), afin d'évaluer le taux de réaction des espèces lorsqu'elles ne sont normalement pas sensées présenter une réaction (figure 3.13).

### 3.2.3. Jeux de données artificiels de taille variable

Pour chacune des 54 courbes de réponse théoriques, le triplet de résumés numériques ( $opt$ ,  $tol$ ,  $pmax$ ) a été transformé en un triplet de coefficients de régression ( $b_0$ ,  $b_1$ ,  $b_2$ ), permettant ainsi la détermination complète de  $p(x)$  le long du gradient de pH (équation 3.6). Nous avons défini 50 valeurs de pH (espacées régulièrement de 0,1 unité de pH) entre 3,1 et 8,0, que l'on a nommées  $x_i$ , et nous avons créé 10 000 données de présence/absence (1/0) pour chaque  $x_i$ , en se basant sur la distribution de Bernoulli associée à la valeur correspondante de  $p(x_i)$ .

Chaque courbe de réponse théorique a donc pu être liée à un tableau  $T$ , de 50 colonnes (une colonne pour chaque valeur de  $x_i$ ) et 10 000 lignes (données de présence/absence), duquel nous avons extrait des jeux de données de taille différente. Par exemple, pour obtenir un jeu de données binaires de 50 relevés soumis à la régression logistique, nous avons sélectionné de manière aléatoire une valeur binaire (un relevé simulé) dans chaque colonne du tableau  $T$ , et nous avons ré-itéré la procédure 100 fois, afin d'obtenir 100 répétitions de taille 50 relevés. De la même manière, pour obtenir un jeu de données de 100 relevés soumis à la régression logistique, nous avons sélectionné de manière aléatoire deux valeurs binaires dans chaque colonne du tableau  $T$ , et nous avons ré-itéré la procédure 100 fois, afin d'obtenir 100 jeux de données de taille 100 relevés. Les 11 tailles différentes de jeux de données que l'on a choisies étaient : 50, 100, 150, 200, 300, 500, 750, 1000, 1500, 2000 et 5000, en sélectionnant de manière aléatoire respectivement 1, 2, 3, 4, 6, 10, 15, 20, 30, 40 et 100 valeurs dans chaque colonne du tableau  $T$ . Pour chaque taille de jeu de données, nous avons produit 100 répétitions différentes.

Pour chacune des trois courbes plates théoriques ( $pmax = 0,05$ ,  $0,20$ , ou  $0,50$ ), 11 jeux de données de taille différente ont été également construits, avec 100 répétitions pour chaque taille.

### 3.2.4. Calcul et évaluation de la qualité des modèles

En tout, nous avons construit 59 400 modèles de régression logistique, puisque chacune des 54 courbes de réponse théoriques ont conduit à 11 tailles de jeux de données (de 50 à 5 000 relevés), qui ont chacune conduit à 100 répétitions. 3 300 autres modèles de régression logistique, correspondant aux trois courbes de réponse théoriquement plates, ont également été construits (trois courbes, 11 tailles de jeux de données et 100 répétitions pour chaque taille). Pour toutes les courbes de réponse calculées, l'optimum (opt) était simplement la valeur de pH pour laquelle la probabilité de présence était maximale (lorsque  $p(x)$  était égal à  $p_{max}$ ). Lorsque la courbe de réponse n'était pas unimodale de manière significative (étant alors soit sigmoïde, soit plate), le calcul de la tolérance (tol), *sensu* ter Braak et Looman (1986), n'était pas possible (équation 3.7). Nous avons donc adopté une autre mesure de l'amplitude écologique (amp), *sensu* Gégout et Pierrat (1998), qui peut s'appliquer pour toutes les formes de réponse et qui représente la gamme de pH dans laquelle les conditions de présence sont optimales. Dans cette étude, comme dans toute la thèse, amp a été calculé comme la gamme de pH contenant 80 % de la distribution de la probabilité de présence.

La qualité des courbes de réponse calculées a été évaluée par la différence entre les valeurs calculées et les valeurs théoriques pour opt, amp, et  $p_{max}$ . De manière subjective, nous avons décidé qu'une courbe de réponse calculée était suffisamment proche de la courbe théorique lorsque (i) la valeur absolue de la différence entre l'optimum calculé et l'optimum théorique ( $|opt_{calculé} - opt_{théorique}|$ ) était strictement inférieure à 1 unité de pH, (ii) la valeur absolue de la différence relative entre l'amplitude calculée et l'amplitude théorique ( $|amp_{calculée} - amp_{théorique}| / amp_{théorique}$ ) était inférieure à 0,25 unité, et (iii) la valeur absolue de la différence relative entre la probabilité maximale calculée et la probabilité maximale théorique ( $|p_{max_{calculée}} - p_{max_{théorique}}| / p_{max_{théorique}}$ ) était inférieure à 0,25 unité.

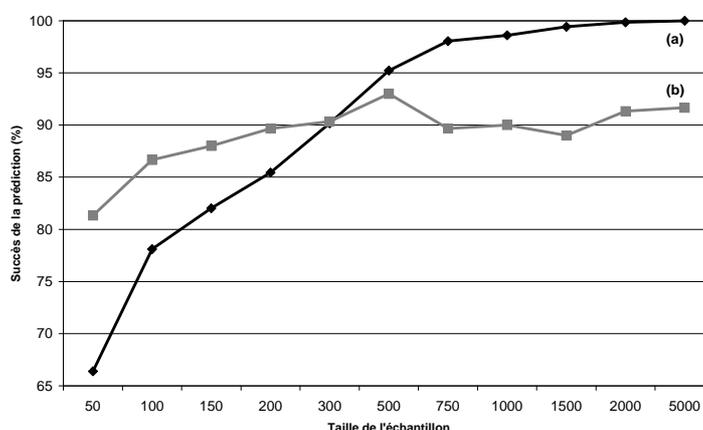
### 3.2.5. Application à des données réelles

Afin d'illustrer l'influence de l'intensité de l'échantillonnage et des caractéristiques des courbes de réponse sur la qualité des modèles calculés, nous avons utilisé des données réelles issues d'EcoPlant, base de données phytoécologiques pour les forêts de France (Gégout 2001, Gégout et al. 2005), qui stocke, pour de nombreux relevés forestiers en France, des informations floristiques (liste des espèces présentes) et des données de milieu (description climatique, description du sol, analyses chimiques d'échantillons de sol, etc.). Par la régression logistique, nous avons caractérisé la réponse écologique des 50 plantes vasculaires les plus fréquentes parmi la végétation forestière spontanée en France. Dans un premier temps, nous avons utilisé un jeu de données de 100 relevés tirés au sort parmi tous les relevés disponibles, puis dans un second temps, nous avons utilisé la totalité des relevés disponibles, soient 3 810 relevés.

### 3.3. Résultats

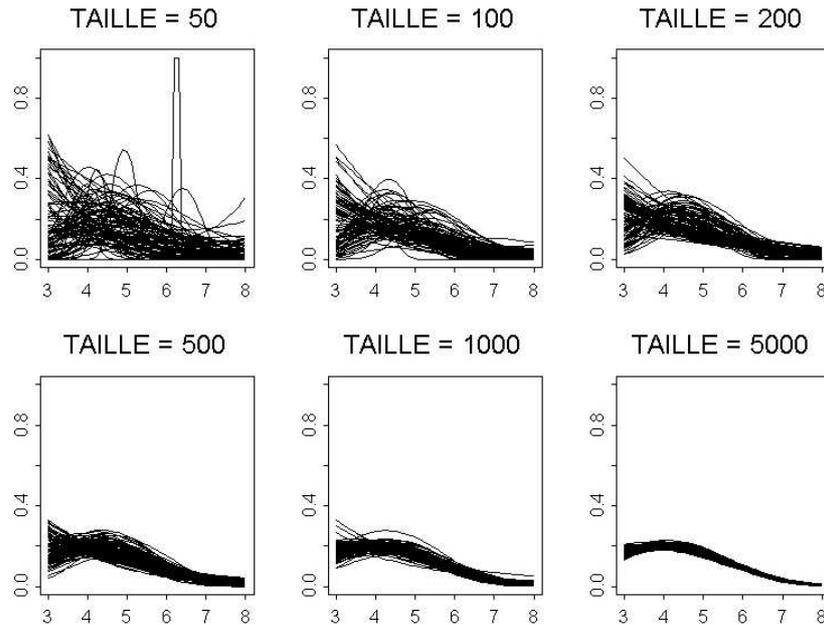
#### 3.3.1. Prédiction de la réponse des espèces au facteur écologique

Parmi les 59 400 modèles calculés pour lesquels une réaction des espèces était attendue, 53 637 modèles (90,3 %) présentait une réaction significative au gradient étudié, les autres 5 763 modèles ne présentant pas de réaction significative (courbe de réponse plate). Parmi les 3 300 modèles calculés pour lesquels une réaction des espèces n'était pas attendue, 2 942 modèles (89,2 %) ne présentait pas de réaction significative, les autres 359 modèles (10,9 %) prédisant une fausse réaction des espèces. Le taux de bonne prédiction de la réaction des espèces, lorsque que la réaction était attendue théoriquement, augmentait avec la taille du jeu de données (figure 3.14a), tandis que le taux de bonne prédiction de l'absence de réaction, lorsque la réaction n'était pas attendue théoriquement, semblait ne pas être affecté par la taille du jeu de données (figure 3.14b).



**Figure 3.14.** Succès de prédiction avec une taille d'échantillon variable (a) du taux de réaction des espèces lorsqu'une réaction est attendue (basé sur 5 400 modèles simulés pour chaque taille d'échantillon), et (b) du taux de non-réaction des espèces lorsqu'une absence de réaction est attendue (basé sur 300 modèles simulés pour chaque taille d'échantillon).

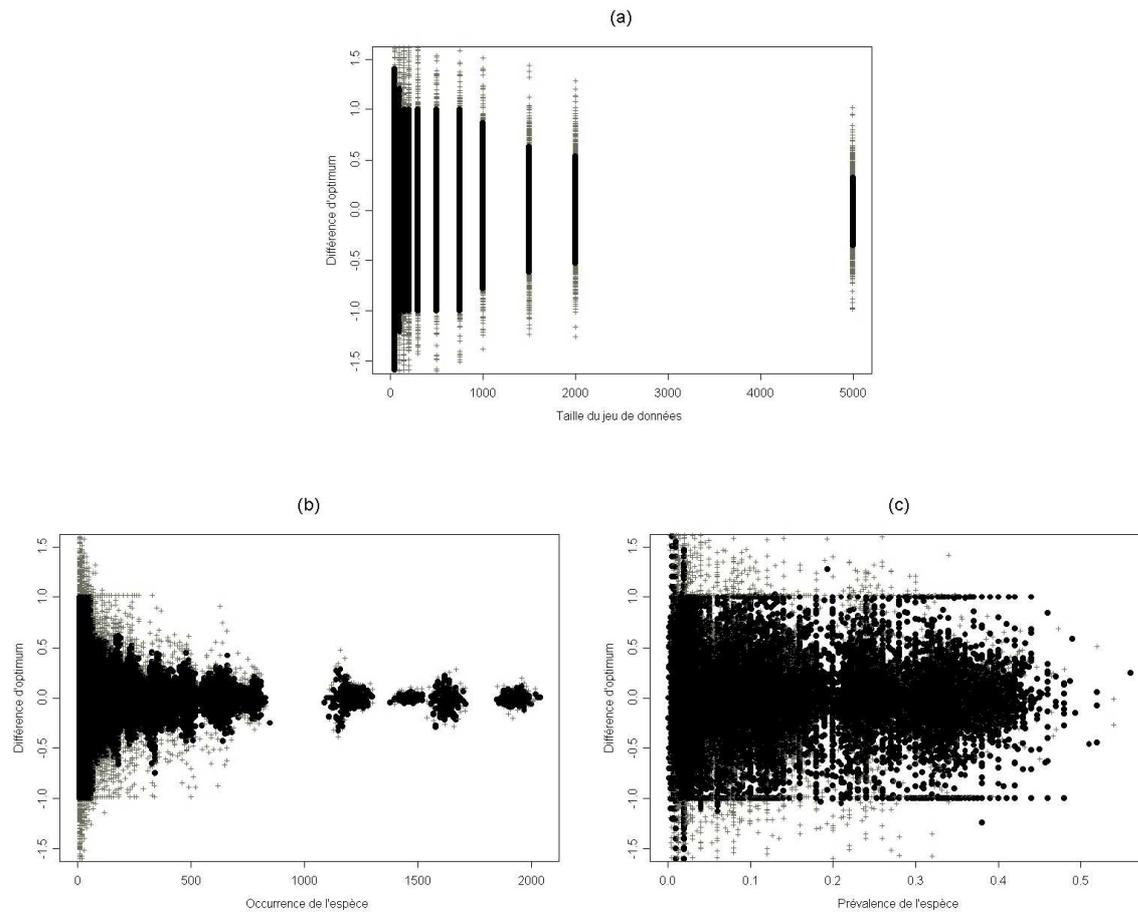
L'influence de la taille du jeu de données sur la qualité du calcul des courbes de réponse était importante (figure 3.15), et révèle que 50 relevés sont insuffisants dans un cas sur trois pour mettre en évidence la réaction significative à un facteur du milieu.



**Figure 3.15.** Courbes de réponse simulées pour la courbe théorique avec  $\text{opt} = 4$ ,  $\text{tol} = 1,5$  et  $\text{pmax} = 0,20$ , et avec une taille variable du jeu de données (50, 100, 200, 500, 1000 et 5000 relevés avec un nombre moyen d'occurrences égal respectivement à 5,8, 11,1, 24,1, 59,0, 118,4 et 595,8). Pour toutes les tailles d'échantillons, le nombre total de modèles simulés est 100 et le nombre de courbes de réponse plates est 59, 33, 4, 0, 0, 0 pour les tailles d'échantillons croissantes, le nombre de modèles sigmoïdes décroissants est 21, 44, 56, 24, 6, 0 et le nombre de courbes de réponse unimodales en forme de cloche est 20, 23, 40, 76, 94, 100.

### 3.3.2. Le nombre d'occurrences est le critère qui permet de sélectionner les espèces à étudier

Pour chacun des 53 637 modèles qui présentaient une réaction attendue,  $\text{opt}$ ,  $\text{amp}$  et  $\text{pmax}$  ont été calculés et comparés aux trois valeurs de la courbe de réponse théorique. La taille du jeu de données et la prévalence des espèces avaient une influence plus faible que le nombre d'occurrences de l'espèce sur la précision d'approximation de l'optimum (figure 3.16). La précision d'approximation de l'optimum réel est meilleure pour des nombres d'occurrences plus élevés (figure 3.16a), tout comme la précision d'approximation de l'amplitude ou de la probabilité maximale de la courbe de réponse (figure 3.17). Un minimum de 50 à 100 occurrences semble être nécessaire pour pouvoir modéliser la réponse d'une espèce au gradient de pH de manière satisfaisante, en termes d'optimum (figure 3.17a), d'amplitude (figure 3.17b), et de probabilité maximale (figure 3.17c).



**Figure 3.16.** Différence, pour les 53 637 modèles présentant une réaction significative au pH, entre l'optimum calculé et l'optimum théorique en fonction de (a) la taille du jeu de données, (b) le nombre d'occurrences de l'espèce, et (c) la prévalence de l'espèce. Pour chaque valeur en abscisses, les points à l'intérieur de l'intervalle comprenant 95 % des valeurs sont représentés par des points noirs, et les points en dehors de cet intervalle (2,5 % de chaque côté) sont représentés par des croix grises.

### 3.3.3. Qualité des modèles calculés et caractéristiques des courbes de réponse

Le pourcentage de bons modèles (valeur absolue de la différence entre l'optimum calculé et l'optimum théorique strictement inférieure à 1 unité de pH, et valeur absolue de la différence relative entre l'amplitude ou la probabilité maximale calculée et théorique inférieure à 0,25 unité), dépend de la taille du jeu de données et des caractéristiques de la courbe de réponse théorique (figure 3.18). Pour les plus petits jeux de données (figure 3.18a), le pourcentage de bons modèles dépendait tout d'abord de  $p_{max}$ , avec de faibles pourcentages pour les faibles valeurs de  $p_{max}$  ; nous n'avons même trouvé aucun bon modèle pour une taille de jeu de données de 100 relevés et une valeur de  $p_{max}$  de 0,05, quelles que soient les valeurs pour  $opt$  et  $amp$ . Pour les plus gros jeux de données (figure 3.18b), le pourcentage de bons modèles était très élevé pour tous les types de courbes de réponse, sauf pour les courbes avec une faible probabilité maximale ( $p_{max} = 0,05$ ).  $p_{max}$  était donc la caractéristique de la courbe ayant la plus forte influence sur le pourcentage de bons modèles.

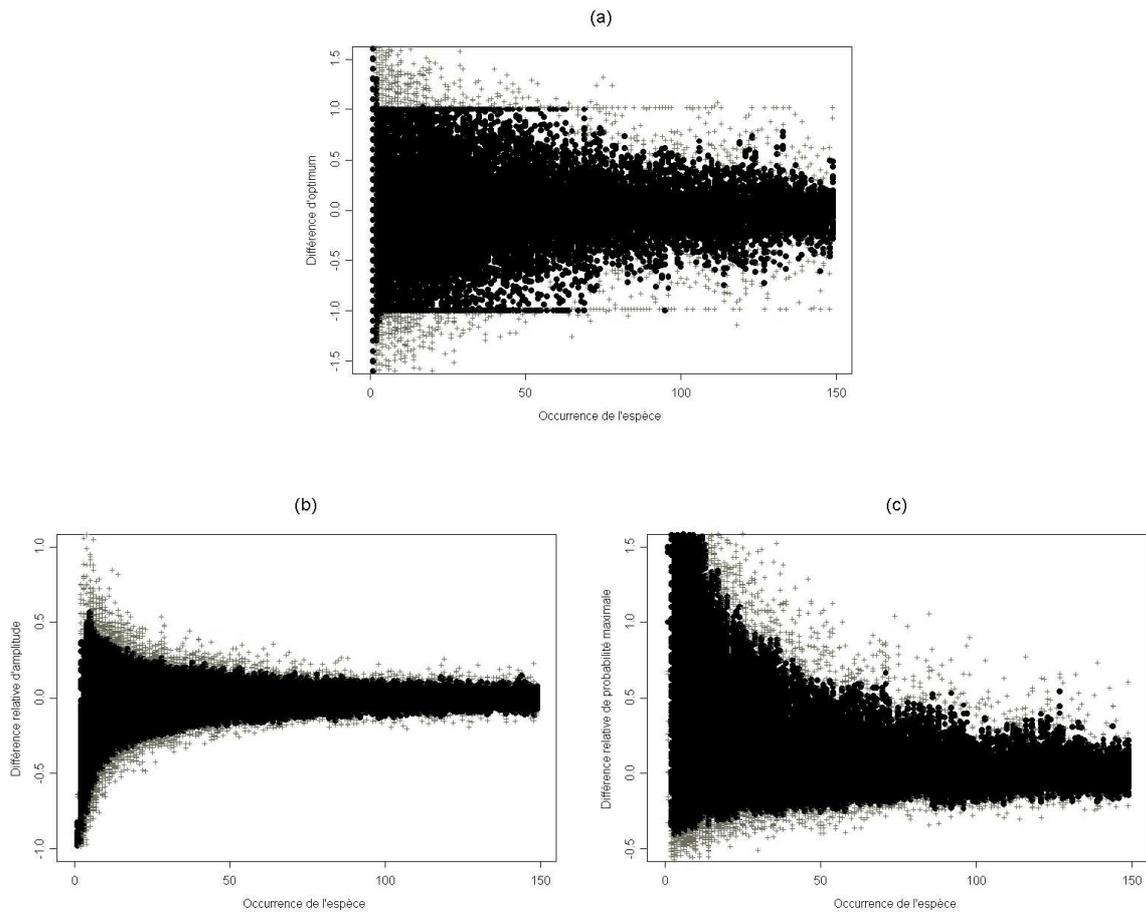


Figure 3.17. Qualité des courbes de réponse simulées pour des espèces à faible occurrence, illustrée par (a) la différence entre l'optimum calculé et l'optimum théorique, (b) la différence relative entre l'amplitude calculée et l'amplitude théorique, et (c) la différence relative entre la probabilité maximale calculée et la probabilité maximale théorique de la courbe.

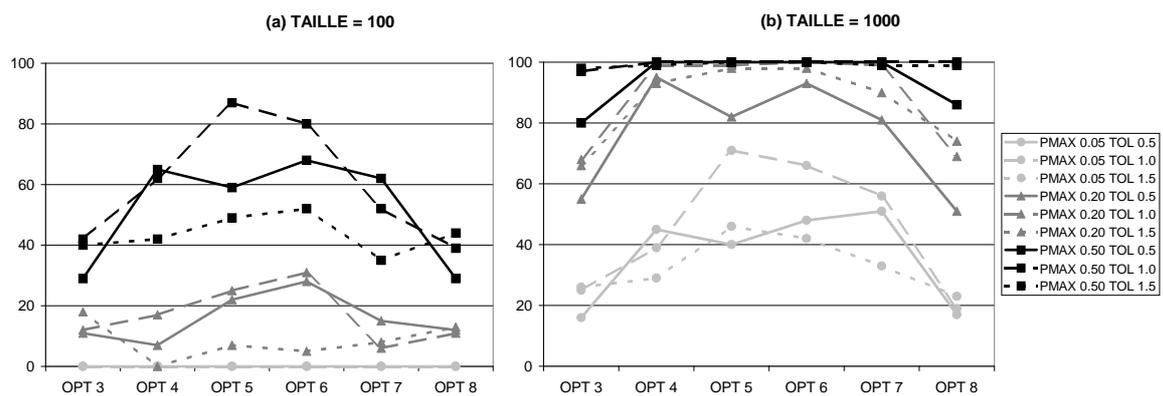


Figure 3.18. Nombre de bons modèles simulés (voir le texte pour la définition) pour deux tailles de jeux de données : (a) 100 relevés, et (b) 1 000 relevés.

La position de l'optimum avait aussi une forte influence sur le pourcentage de bons modèles, avec de meilleurs modèles pour des valeurs intermédiaires de l'optimum. En effet, lorsque l'optimum théorique d'une espèce se trouve près d'une extrémité du gradient, les occurrences de l'espèce ne sont rencontrées que dans une portion limitée du gradient et ne sont souvent pas suffisantes pour conduire à une courbe de réponse unimodale de manière significative, mais conduisent plutôt à une courbe de réponse apparemment sigmoïde (voir la figure 3.15). Par exemple, parmi les 9 900 modèles calculés avec un optimum théorique égal à 4, 7 519 modèles ont conduit à une courbe de réponse unimodale, tandis que 1 542 et 839 modèles ont conduit respectivement à une courbe sigmoïde décroissante et plate. Par contre, lorsque l'optimum théorique d'une espèce se trouve au centre du gradient étudié, les occurrences de l'espèce sont rencontrées le long d'une plus grande portion du gradient et sont souvent suffisantes pour conduire à une courbe de réponse unimodale significative.

L'influence de l'amplitude sur le pourcentage de bons modèles était importante pour les espèces avec un optimum théorique de valeur intermédiaire. Par exemple, parmi les 9 900 modèles calculés avec un optimum théorique de 5, 8 303 modèles ont conduit à une courbe unimodale, tandis que 200 et 1 397 modèles ont respectivement conduit à une courbe sigmoïde décroissante et plate. Les plus grandes amplitudes ( $tol = 1,5$ ) couplées aux valeurs intermédiaires d'optimum ont conduit à des occurrences d'espèces réparties de manière régulière tout le long du gradient, et ainsi conduit à plus de courbes de réponse apparemment plates.

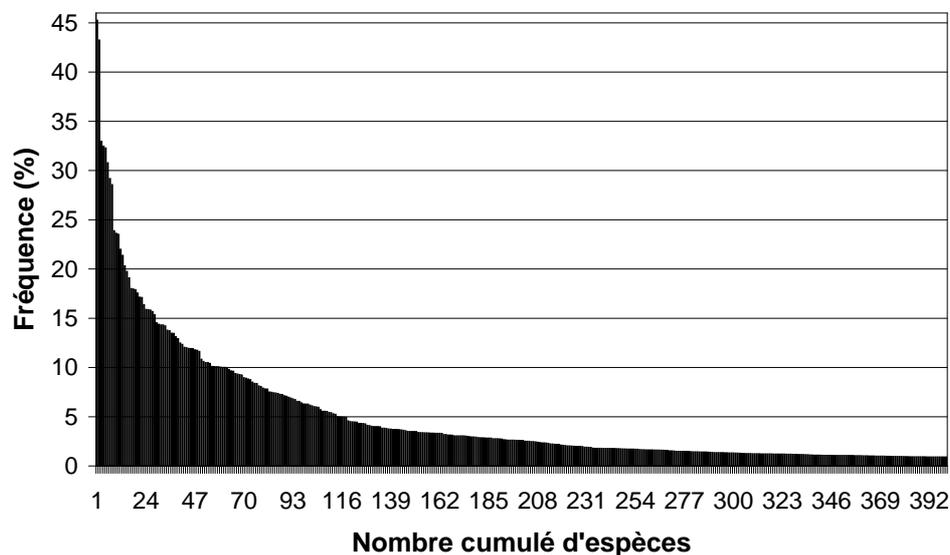
Les espèces les plus faciles à modéliser, celles qui nécessitent un jeu de données de petite taille (tableau 3.4), étaient des espèces fréquentes ( $p_{max} = 0,50$ ), avec des positions intermédiaires pour l'optimum ( $opt = 4, 5, 6, \text{ or } 7$ ), et des faibles valeurs d'amplitude ( $tol = 0,5$  ou  $1,0$ ) ; tandis que les espèces les plus difficiles à modéliser, celles qui nécessitent un jeu de données de grande taille (tableau 3.4), étaient des espèces peu fréquentes ( $p_{max} = 0,05$ ), avec l'optimum proche d'une extrémité du gradient ( $opt = 3$  ou  $8$ ) et une amplitude large ( $tol = 1,5$ ).

**Tableau 3.4.** Taille nécessaire du jeu de données (nombre de relevés) pour obtenir 95 % de bons modèles (voir le texte pour la définition, et la figure 3.18).

<b>pmax</b>	<b>tol</b>	<b>opt 3</b>	<b>opt 4</b>	<b>opt 5</b>	<b>opt 6</b>	<b>opt 7</b>	<b>opt 8</b>
<b>0.05</b>	<b>0.5</b>	> 5000	> 5000	5000	> 5000	> 5000	> 5000
<b>0.05</b>	<b>1.0</b>	> 5000	5000	5000	5000	5000	> 5000
<b>0.05</b>	<b>1.5</b>	> 5000	5000	5000	5000	5000	> 5000
<b>0.20</b>	<b>0.5</b>	5000	1000	1500	1500	2000	5000
<b>0.20</b>	<b>1.0</b>	5000	1000	750	1000	1000	5000
<b>0.20</b>	<b>1.5</b>	5000	1500	1000	750	1500	5000
<b>0.50</b>	<b>0.5</b>	2000	300	300	300	300	2000
<b>0.50</b>	<b>1.0</b>	1000	300	150	150	300	1000
<b>0.50</b>	<b>1.5</b>	1000	750	500	300	750	1000

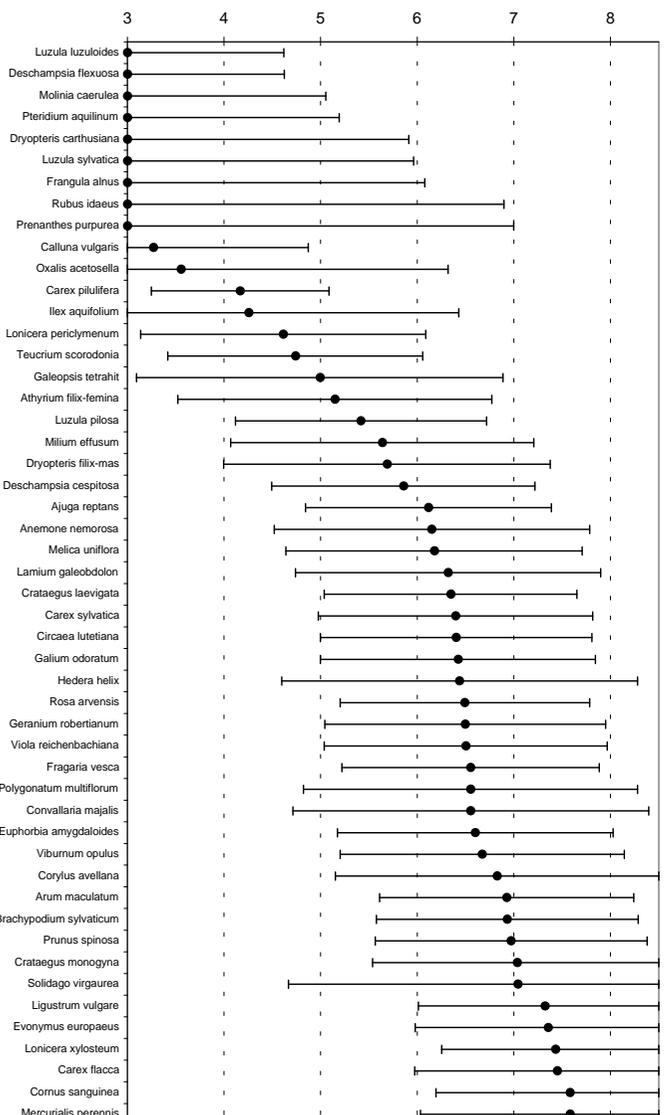
### 3.3.4. Application à des données réelles

La base de données phytoécologiques EcoPlant illustre le fait que dans la nature, il y a peu d'espèces fréquentes et beaucoup d'espèces peu fréquentes (figure 3.19). Nous n'avons présenté, sur la figure 3.19, que les 400 plantes forestières les plus fréquentes intégrées à EcoPlant, parmi 2 154 taxa (Gégout et al. 2005). Pour les 50 plantes vasculaires les plus fréquentes dans les forêts de France (en excluant les arbres), une courbe de réponse écologique a été calculée le long du gradient de pH (entre 3,0 et 8,5), et ces 50 espèces ont été ordonnées selon leur optimum de pH calculé croissant (figure 3.20). Pour le premier jeu de données, tous les 3 810 relevés disponibles avec une mesure du pH eau du sol et répartis sur toute la France, ont été utilisés, le nombre minimal d'occurrences étant 321 pour *Ajuga reptans*. Les 50 courbes de réponse calculées devaient donc être une représentation raisonnable du comportement réel des espèces vis-à-vis du pH du sol en France (figure 3.20a). Pour le second jeu de données (100 relevés tirés au hasard parmi les 3 810 précédents), le nombre minimal d'occurrences était 4 pour *Galeopsis tetrahit*. Le calcul des courbes de réponse pour ces 50 mêmes espèces, mais avec seulement 100 relevés, illustre deux biais causés par la régression logistique (figure 3.20b) : (i) une tendance à prédire une absence de réaction pour de nombreuses espèces avec une valeur intermédiaire d'optimum de pH et une large amplitude (par exemple *Galeopsis tetrahit*, *Athyrium filix-femina*, *Milium effusum*, *Dryopteris filix-mas*), et (ii) une tendance à prédire un optimum extrême pour certaines espèces dont l'optimum réel se trouve près de l'extrémité du gradient, mais pas à l'extrémité (par exemple *Calluna vulgaris*, *Lonicera periclymenum* pour l'extrémité inférieure du gradient de pH et par exemple *Ligustrum vulgare*, *Cornus sanguinea*, *Carex flacca*, *Evonymus europaeus*, *Ligustrum vulgare*, pour l'extrémité supérieure du gradient de pH).



**Figure 3.19.** Fréquence décroissante (%) des 400 plantes forestières les plus fréquentes intégrées à EcoPlant, base de données phytoécologiques pour les forêts de France (Gégout et al. 2005). La figure est basée sur 6 432 relevés floristiques complets.

(a) TAILLE = 3 810



(b) TAILLE = 100

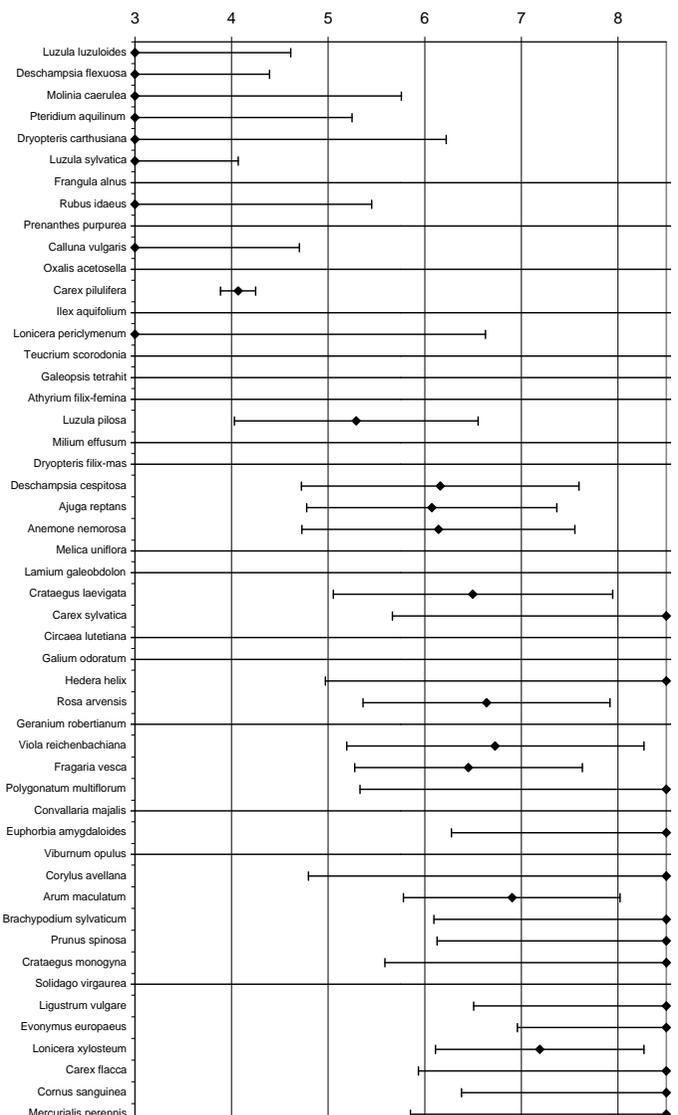


Figure 3.20. Optimum et amplitude écologiques des 50 plantes les plus fréquentes présentes dans EcoPlant, base de données phytécologiques pour les forêts de France, calculés sur la base de (a) 3 810 relevés, et (b) 100 relevés.

### 3.4. Discussion

La taille du jeu de données est une conséquence directe de la phase d'échantillonnage de l'étude et se décide souvent avant d'effectuer le moindre relevé. La décision d'un nombre de relevés détermine ainsi le nombre d'espèces qui peuvent être modélisées. Nos résultats ont montré que le nombre d'occurrences des espèces était un facteur important pour calculer des courbes de réponse écologiques raisonnables par la régression logistique, et nous proposons une valeur minimale de l'ordre de 50 à 100 occurrences pour obtenir des résultats satisfaisants. En effet, la régression logistique a semblé être relativement faible pour déterminer de manière précise les caractéristiques des courbes des espèces peu fréquentes, et [ter Braak et Looman \(1986\)](#) ont montré que même la technique des moyennes pondérées pouvait fournir une estimation raisonnable de l'optimum d'espèces peu fréquentes.

Il est important de connaître le seuil minimal d'occurrences nécessaire à l'obtention de courbes de réponse réalistes pour sélectionner les espèces qui peuvent être étudiées au sein d'une flore ([Wiser et al. 1998](#), [Pearce et Ferrier 2000a, 2000b](#), [Elith et Burgman 2002](#)). La valeur de ce seuil est essentielle car la plupart des plantes forestières recensées dans la base de données EcoPlant sont rencontrées dans moins de 50 relevés ([Gégout et al. 2005](#)) et baisser la valeur du seuil pourrait permettre de traiter plus d'espèces ([Stockwell et Peterson 2002a](#)). Cependant, un nombre d'occurrences de plus en plus élevé ne permettra pas d'augmenter de manière significative de la précision de détermination des caractéristiques de la courbe, et une valeur maximale comprise en 500 et 1 000 occurrences pour une espèce peut également être suggérée ([Virtanen et al. 1998](#), [Stockwell et Peterson 2002b](#)).

Sur la base à la fois de données artificielles et de données réelles, nous avons illustré le fait que certaines espèces avec un optimum théorique au centre du gradient, pouvaient conduire à des courbes de réponse faussement plates lorsque l'amplitude est trop large, réduisant ainsi le nombre d'espèces avec un optimum apparemment intermédiaire (figure 3.20). De la même manière, certaines espèces avec un optimum réel proche d'une extrémité du gradient pouvaient conduire à des courbes de réponse faussement sigmoïdes. Une solution technique pour résoudre ce problème pourrait être de conserver le terme quadratique dans les équations de régression logistique, chaque fois que le terme linéaire est significatif. Une solution pratique serait d'augmenter l'intensité de l'échantillonnage dans des zones proches des extrémités du gradient, ce qui confirme des observations précoces de [Mohler \(1983\)](#).

Un travail de recherche important a été effectué sur la forme des courbes de réponse, pour déterminer si les espèces présentent plutôt des réponses symétriques ou asymétriques ([Austin et Nicholls 1997](#), [Oksanen 1997](#), [Rydgren et al. 2003](#)), mais l'intérêt de cette étude était de vérifier si une courbe de réponse écologique connue pouvait être reconstruite, selon des tailles de jeux de données et des caractéristiques de courbes de réponse différents. Nous pensons que le comportement écologique des espèces pourrait être analysé plus en profondeur si la taille des jeux de données utilisés est augmentée.



## Deuxième partie

Stabilité du comportement écologique des espèces à l'échelle de la France



#### 4. Comparaison de la réponse écologique des espèces forestières herbacées vis-à-vis du pH du sol dans le nord-est et le nord-ouest de la France

*Article associé (Annexe 2)*

**Coudun, Ch. et J.-C. Gégout (2005)** Ecological behaviour of herbaceous forest species along the pH gradient: a comparison between oceanic and semi-continental regions in northern France. *Global Ecology & Biogeography*, **14**, 263-270.

## Résumé

En se basant sur 2402 relevés phytotécologiques disponibles avec une liste complète d'espèces et une valeur mesurée de pH, issue d'analyse chimique du premier horizon minéral du sol, cette étude compare de manière quantitative la réponse écologique de 46 espèces forestières herbacées le long du gradient pH dans deux régions du nord de la France : le nord-ouest océanique (NW) et le nord-est semi-continentale (NE).

Pour chacune des 46 espèces qui présentent plus de 50 occurrences dans chacune des deux régions NW et NE, nous avons calculé une courbe de réponse écologique par des modèles de régression logistique simple, et deux valeurs numériques synthétiques ont pu être dérivées : l'optimum écologique (OPT) et l'amplitude écologique (AMP). La comparaison du comportement écologique des espèces présentes dans les deux régions a été effectuée en terme de déplacement de l'optimum et/ou de l'amplitude.

Les 46 espèces étudiées ont toutes révélé une réaction significative au pH. Notre observation principale a été la relative stabilité du comportement de la plupart des espèces vis-à-vis des conditions de pH, ce qui est cohérent avec les résultats d'études antérieures. Des petites différences entre les deux régions, concernant les deux paramètres synthétiques (OPT et AMP) sont cependant soulignées, constituant vraisemblablement une forme de libération de niche due à la compétition ("competitive release") pour les conditions neutres de pH dans la région NW.

Parmi les principales explications d'un changement géographique du comportement écologique des espèces herbacées, nous suggérons que les effets de la compétition soient l'explication la plus plausible dans notre étude. A la lumière des études antérieures, il semble que des analyses plus approfondies des changements géographiques de comportement écologique des espèces végétales forestières soient nécessaires sur de plus vastes zones, comme par exemple le continent européen.

### *Mots-clés*

Niche écologique, amplitude écologique, optimum écologique, courbes de réponse écologique, base de données EcoPlant, valeurs indicatrices d'Ellenberg, espèces forestières herbacées, France, régression logistique.

### *Nomenclature*

[Tutin et al. \(2001\)](#).

## 4.1. Introduction

La modélisation de la niche écologique des espèces végétales a reçu beaucoup d'attention dans les dernières décennies, pour quantifier la réponse des espèces aux facteurs abiotiques. De nombreuses techniques statistiques ont été développées pour dessiner et interpréter des courbes de réponse et permettre ainsi une compréhension plus profonde des relations espèces/environnement (Guisan et Zimmermann 2000), et les modèles linéaires généralisés (GLM) ou les modèles additifs généralisés (GAM) constituent les techniques les plus intensément utilisées (par exemple Austin et al. 1984, Odland et al. 1995). Récemment, des colloques internationaux ont conduit à de nombreuses publications concernant la modélisation et la prédiction de la distribution des espèces, avec des développements à la fois théoriques et pratiques (Guisan et al. 2002, Lehmann et al. 2002a, Scott et al. 2002).

L'établissement de courbes de réponse unimodales permet de définir des valeurs indicatrices qui présentent un grand intérêt pour synthétiser l'information écologique des espèces (ter Braak et Looman 1986, Lawesson et Oksanen 2002). L'optimum et l'amplitude écologiques sont traditionnellement issus de calculs numériques sur la courbe de réponse (ter Braak et Gremmen 1987, Odland et al. 1995) et facilitent ainsi la comparaison de la réaction des espèces aux facteurs écologiques. Des valeurs indicatrices empiriques ont été proposées par Ellenberg et al. (1992) pour l'Europe centrale continentale, et ces valeurs ont été utilisées avec succès dans de nombreuses autres régions (Diekmann 2003). Cependant, les valeurs indicatrices d'Ellenberg ont été revisitées pour tenir compte des conditions au niveau local, comme par exemple aux Pays-Bas (Ertsen et al. 1998), au Danemark (Lawesson et Mark 2000), en Suède (Diekmann 1995) ou en Grande-Bretagne océanique (Hill et al. 1999, 2000). En plus du fait qu'elle puisse varier dans l'espace, la valeur indicatrice d'une plante particulière peut également varier avec le temps, au cours de la vie de la plante (Parrish et Bazzaz 1985).

Depuis le début des années 1990, de nombreuses études ont consisté en la comparaison de la réponse écologique des plantes dans des régions différentes (Thompson et al. 1993, Diekmann et Lawesson 1999, Hill et al. 2000, Prinzing et al. 2002, Gégout et Krizova 2003). Diekmann et Lawesson (1999) ont suggéré trois explications principales pour un changement géographique possible du comportement écologique : (i) les espèces peuvent se différencier en écotypes qui sont adaptés à des régions différentes (Turesson 1922), (ii) la loi de Walter, connue aussi comme "loi de constance de l'habitat" (Walter et Walter 1953), peut forcer une espèce à compenser un facteur écologique par un autre et ainsi modifier l'optimum ou l'amplitude écologique de l'espèce le long d'un gradient, ou (iii) la présence/absence de compétiteurs peut causer une variation de la valeur indicatrice d'une espèce. En France, des caractéristiques indicatrices empiriques ont été proposées par Rameau et al. (1989, 1993) pour les espèces végétales forestières, et au début des années 1970, Le Tacon et Timbal (1972, 1973) ont relevé des différences de comportement écologique de certaines espèces végétales forestières dans des peuplements de hêtre dans le nord-est et le nord-ouest de la France, vis-à-vis de la nature de l'humus du sol.

Diekmann et Lawesson (1999) ont noté que le pH du sol était l'une des variables principales ayant une influence sur la variation de la flore au sein des forêts de leurs quatre zones d'étude, et ils ont également noté que des mesures de terrain ou des expériences "faisaient cruellement défaut". Trop souvent, la valeur des facteurs du milieu est déterminée à partir de la moyenne pondérée des valeurs indicatrices des espèces pour les facteurs proposés par Ellenberg et al. (1992) (voir par exemple Diekmann 1996, Dzwonko 2001). Le pH représente un gradient complexe, mais il a été considéré comme étant un gradient écologique direct (Austin 1980) parce qu'il a un effet physiologique direct sur la croissance des végétaux. On peut aussi le considérer comme un gradient de ressource puisqu'il contrôle le prélèvement des plantes en éléments minéraux et il est,

dans le même temps, corrélé à de nombreux facteurs édaphiques et climatiques (Duchaufour 1989, Falkengren-Grerup et al. 1995, Tyler 2003).

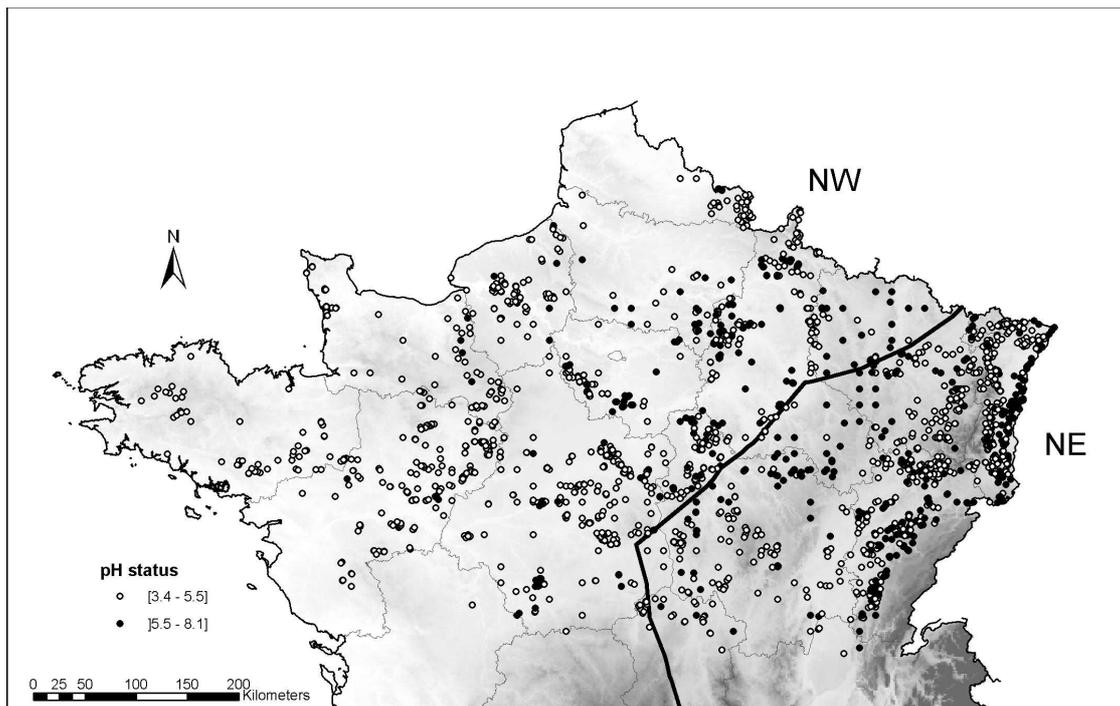
Dans ce chapitre, l'objectif principal était d'examiner de manière quantitative la réponse écologique des plantes forestières vis-à-vis du pH du sol dans deux régions du nord de la France (le nord-est semi-continentale, NE et le nord-ouest océanique, NW), en se basant sur des mesures réelles du pH du sol issues d'analyses chimiques du premier horizon minéral du sol forestier. Cela constitue une différence majeure avec les études précédentes qui basaient leurs calculs sur les valeurs indicatrices empiriques d'Ellenberg. En travaillant avec deux régions, le NW océanique et le NE semi-continentale, nous voulions observer si les espèces forestières herbacées présentent des courbes de réponse écologique similaires dans des régions aux conditions écologiques plutôt similaires, mais séparées géographiquement.

## 4.2. Matériels et méthodes

### 4.2.1. Deux jeux de données régionaux

Deux jeux de données régionaux ont été extraits d'EcoPlant (Gégout 2001, Gégout et al. 2005), une base de données phytoécologiques intégrant aujourd'hui plus de 110 sources différentes de relevés, disséminés sur tout le territoire français métropolitain et la Corse. L'objectif principal d'EcoPlant est de stocker des relevés floristiques complets avec toute l'information écologique disponible associée (paramètres liés au climat et au sol), soit mesurée sur le terrain, soit analysée chimiquement en laboratoire ou soit issue de croisements sous systèmes d'information géographique. EcoPlant est actuellement utilisée pour calculer des valeurs indicatrices écologiques analytiques pour les espèces végétales forestières, vis-à-vis de facteurs climatiques et édaphiques (Gégout 2001, Gégout et al. 2005).

Dans cette étude, 73 sources de relevés ont été utilisées (la liste des références bibliographiques est disponible auprès des auteurs) et deux régions du nord de la France ont été définies, en se basant sur une distance de 300 km à la mer (nord-ouest : NW et nord-est : NE, voir la figure 4.21). Nous avons séparé les relevés océaniques des relevés semi-continentaux pour être cohérents avec des cartes bioclimatiques de France existantes (Bessemoulin 1989). Nous obtenons ainsi deux jeux de données régionaux similaires en termes de nombres de relevés (1 353 relevés forestiers pour le NW et 1 049 pour le NE). Afin de garder les composantes écologiques majeures, sauf la nutrition minérale (le pH du sol) les plus constantes possibles, les effets de la température ont été neutralisés dans la phase d'échantillonnage de l'étude. La gamme de températures moyennes annuelles a été fixée entre 8,5 et 11,5 °C et tous les relevés choisis ont une altitude inférieure à 600 m. L'effet de la lumière a été aussi neutralisé de manière implicite puisque tous les relevés proviennent de milieux forestiers fermés.



**Figure 4.21.** Localisation des 2 402 relevés forestiers étudiés dans les deux régions du nord de la France (nord-ouest : NW et nord-est : NE). La classe de pH est représentée à travers la couleur du symbole.

Notre jeu de données couvrait la totalité du gradient de pH dans les deux régions, puisque le pH du sol est compris entre 3,4 et 8,1. Toutes les valeurs de pH sont des valeurs de pH eau mesurées en laboratoire pour l'horizon organo-minéral A du sol. Nous avons utilisé une séparation pragmatique du gradient de pH en deux classes, pour tenir compte des stations forestières acides ( $3,4 \leq \text{pH} \leq 5,5$ ) et neutres à riches ( $5,5 < \text{pH} \leq 8,1$ ) (voir par exemple [Gough et al. 2000](#)). La valeur 5,5 pour le pH représente également une valeur sous laquelle une toxicité aluminique peut être importante pour les plantes ([Espiau et Peyronel 1977](#), [Badeau 1998](#)). Parmi les 1 353 stations forestières de la région NW, 1 083 étaient acides et 270 étaient neutres à riches selon la description précédente, et parmi les 1 049 stations forestières de la région NE, 668 étaient acides et 381 étaient neutres à riches.

La plupart des relevés forestiers échantillonnés ("quadrats") présentaient une aire de 400 m<sup>2</sup>, en cohérence avec la pratique phytoécologique actuelle, et des relevés floristiques complets étaient disponibles pour chaque site (information en présence/absence). Pour pouvoir être intégrées dans des analyses plus approfondies, les espèces forestières herbacées devaient être présentes au moins 50 fois dans chaque région (c'est-à-dire présentes dans au moins 5 % des relevés de chaque région). En tout, le pool d'espèces étudiées dans ce chapitre représente 46 espèces herbacées.

#### **4.2.2. Modélisation par régression logistique simple**

La réponse écologique des espèces au pH du sol a été calculée par des modèles de régression logistique simple et l'établissement de courbes de réponse ([ter Braak et Looman 1986](#)). La régression logistique est une technique qui fait partie des modèles linéaires généralisés ([McCullagh et Nelder 1997](#)), avec une fonction de lien logistique et une distribution binomiale, et qui est souvent utilisée pour caractériser la relation entre les espèces et leur environnement ([Austin et al. 1984](#), [ter Braak et Looman 1986](#), [Odland et al. 1995](#)). Nous avons caractérisé la significativité (au seuil de 5 %,  $p = 0,05$ ) du modèle logistique gaussien (courbe de réponse unimodale en forme de cloche), face au modèle logistique linéaire (courbe de réponse sigmoïde croissante ou décroissante), ou face au modèle nul (pas de réaction et courbe de réponse plate) par un test sur

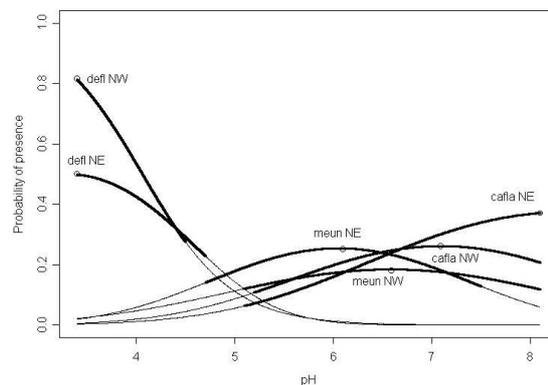
la déviance résiduelle (McCullagh et Nelder 1997) basé sur le critère d'information d'Akaike (Akaike 1973). Tous les calculs ont été effectués à l'aide du logiciel statistique S-Plus (MathSoft 1999).

La technique de modélisation par régression logistique reste cependant soumise au débat, puisqu'elle ne permet pas de modéliser toutes les formes de courbes de réponse (Austin et al. 1984, Austin 1987, 2002, Austin et Nicholls 1997, Oksanen 1997, Lawesson et Oksanen 2002). Nous avons contourné ce problème en calculant deux paramètres qui résument l'information des courbes de réponse dérivées pour chacune des 46 espèces herbacées étudiées : l'optimum écologique (OPT) et l'amplitude écologique (AMP). OPT est simplement la valeur du pH pour laquelle la probabilité de présence atteint son maximum, et lorsque cette valeur était comprise entre 3,4 et 8,1 (réponse unimodale significative), nous avons pu calculer un intervalle de confiance autour de cet optimum (ter Braak et Looman 1986, Vetaas 2000, Oksanen et al. 2001). AMP représente la gamme de pH dans laquelle les conditions de présence sont optimales et a été calculée comme la gamme de pH contenant 80 % de la distribution de la probabilité de présence (Gégout et Pierrat 1998). Une telle mesure de l'amplitude permet une comparaison claire entre les deux régions et offre l'avantage d'être applicable pour toute forme de réponse, contrairement à la mesure de tolérance de ter Braak et Looman (1986), qui n'est valable que pour les modèles logistiques gaussiens.

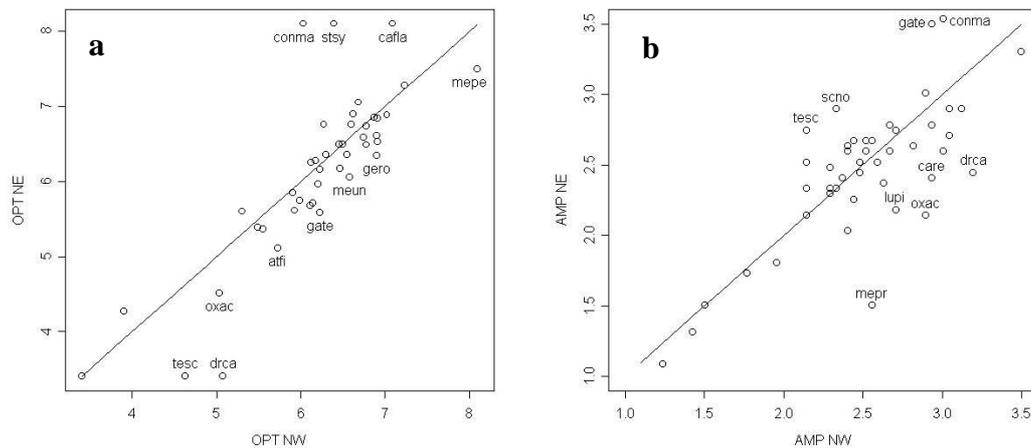
### 4.3. Résultats

#### 4.3.1. Stabilité générale de la réponse au pH du sol pour la plupart des espèces

Dans les deux régions, les 46 espèces végétales étudiées ont révélé une réaction significative au pH, justifiant le choix du pH comme variable écologique importante. La figure 4.22 montre des courbes de réponse pour les espèces acidiphile *Deschampsia flexuosa*, neutrophile *Melica uniflora* et calcicole *Carex flacca*, issues de modèles logistiques dans la deux régions. Les optima et les amplitudes des 46 espèces herbacées sont présentés dans la figure 4.23. Nous avons trouvé que les espèces tendent à occuper les mêmes positions de niche dans le NW et le NE : le coefficient de détermination  $R^2$ , calculé pour les 46 espèces pour caractériser le lien entre les valeurs d'OPT entre le NW et le NE, vaut 0,78, et le coefficient de détermination vaut 0,57 pour les valeurs d'AMP entre le NW et le NE.



**Figure 4.22.** Courbes de réponse écologique vis-à-vis du pH, calculées par régression logistique, dans le NW et le NE pour *Deschampsia flexuosa* (defl), *Melica uniflora* (meun) et *Carex flacca* (cafla). Les points indiquent la position de l'optimum de pH (OPT) et les lignes grasses indiquent la largeur de l'amplitude écologique (AMP).

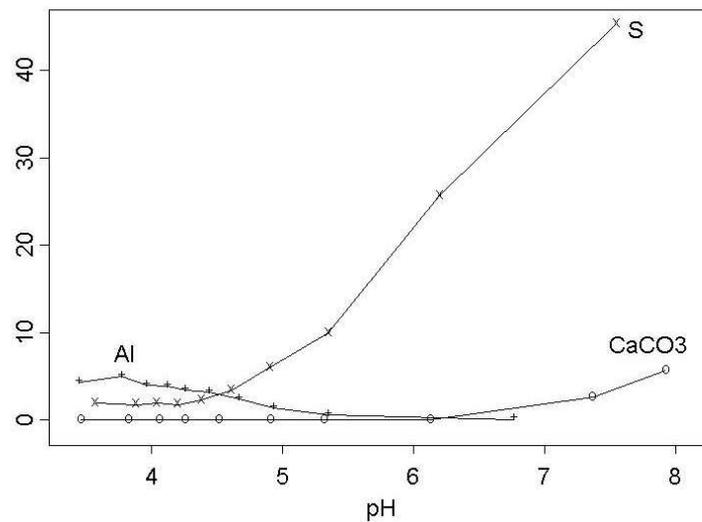


**Figure 4.23.** Représentation des positions (a) des optima de pH (OPT) et (b) des amplitudes de pH dans le NW et le NE pour les 46 espèces forestières herbacées pour lesquelles une courbe de réponse écologique a pu être ajustée dans les deux régions (voir le tableau 4.5 pour les noms complets des espèces qui sont identifiées ici par leur seul radical).

La répartition d'espèces apparemment acidiphiles ( $OPT \leq 5.5$ ) et neutrophiles ( $OPT > 5.5$ ) dans les deux régions est comme suit : la plupart des espèces (43 sur 46) appartiennent au même groupe écologique vis-à-vis du pH dans les deux régions, avec neuf espèces acidiphiles et 34 neutrophiles. Seule *Rubus fruticosus*, une espèce qui intègre probablement différents taxa dans les deux régions, semble être acidiphile dans le NW et neutrophile dans le NE. A l'inverse, *Athyrium filix-femina* et *Stellaria holostea* semblent être neutrophiles dans le NW et acidiphiles dans le NE. Toutefois, l'amplitude écologique observée pour ces trois espèces est relativement large.

#### 4.3.2. La compétition pour les environnements les plus favorables

La densité des optima des espèces est maximale pour des valeurs de pH du sol comprises entre 6 et 7, ce qui représente une gamme dans laquelle les conditions de nutrition minérale sont optimales dans les forêts du nord de la France. Pour de telles valeurs de pH comprises entre 6 et 7, la toxicité due à l'aluminium échangeable ne s'observe plus et la toxicité due au calcaire actif ( $\text{CaCO}_3$ ) pour des valeurs de pH supérieures, ne s'observe pas encore (figure 4.24). Dans ces conditions de pH, la somme des cations échangeables (calcium, magnésium et potassium) est aussi favorable et contribue à des conditions optimales de nutrition minérale pour lesquelles les espèces peuvent entrer en compétition.



**Figure 4.24.** Teneur moyenne du premier horizon minéral du sol (horizon A) en aluminium échangeable (Al, en meq par 100 g de sol sec), en somme de cations nutritifs échangeables (S, en meq par 100 g de sol sec), et en calcaire actif (CaCO<sub>3</sub>, en %) le long du gradient de pH, dans le nord de la France. Les calculs ont été effectués sur la base de 2 798 relevés forestiers pour S, 1 208 relevés pour Al et 333 relevés pour CaCO<sub>3</sub> (source : EcoPlant, Gégout et al. 2005).

On peut observer une tendance pour certaines espèces d'être acido-tolérantes à neutrophiles dans le NE et de présenter un optimum de pH plus élevé dans le NW (c'est le cas par exemple de *Oxalis acetosella*, *Athyrium filix-femina*, *Galeopsis tetrahit*, *Melica uniflora* ou *Geranium robertianum*), tandis que d'autres espèces qui sont calcicoles dans le NE présentent un optimum de pH inférieur dans le NW (c'est le cas par exemple de *Carex flacca*, *Stachys sylvatica* ou *Convallaria majalis*, voir la figure 4.23a). Toutes les espèces qui présentent un changement d'optimum et/ou d'amplitude supérieur à 0,5 unité de pH sont listées dans le tableau 4.5. De manière générale, il apparaît que les espèces présentent des optima de pH moins extrêmes dans le NW que dans le NE (figure 4.23a).

Le nombre d'espèces par relevé est supérieur dans les forêts du NE par rapport à celles du NW, pour toutes les valeurs de pH, et plus particulièrement pour des valeurs optimales comprises entre 6 et 7 (figure 4.25). De plus, le pool spécifique régional d'espèces neutrophiles/calcicoles semble être plus large dans le NE que dans le NW. En effet, 19 espèces supplémentaires peuvent être rencontrées plus de 50 fois dans la région NE et parmi elles, 12 étaient calcicoles. Parmi les 13 espèces présentes dans plus de 50 relevés du NW mais pas dans le NE, seules cinq espèces étaient calcicoles. Avec moins d'espèces par relevé sur des sols eutrophes et avec un pool d'espèces neutrophiles plus petit dans la région NW, nous suggérons que les effets de compétition sont moins marqués dans le NW que dans le NE, et ceci peut expliquer la tendance observée de valeurs d'optimum de pH dans les meilleures conditions dans le NW. Des conditions de compétition supposées plus faibles dans le NW que dans le NE peuvent alors expliquer le glissement d'espèces acido-tolérantes et calcicoles dans le NE vers des optima neutres dans le NW.

**Tableau 4.5.** Liste des espèces forestières herbacées qui présentent un changement d'optimum ou d'amplitude de pH supérieur à 0,5 unité entre les deux régions d'étude NW et NE. OPT représente l'optimum écologique, CI95 représente l'intervalle de confiance (95 %) autour de l'optimum et AMP représente l'amplitude.

**Changement pour OPT et AMP**

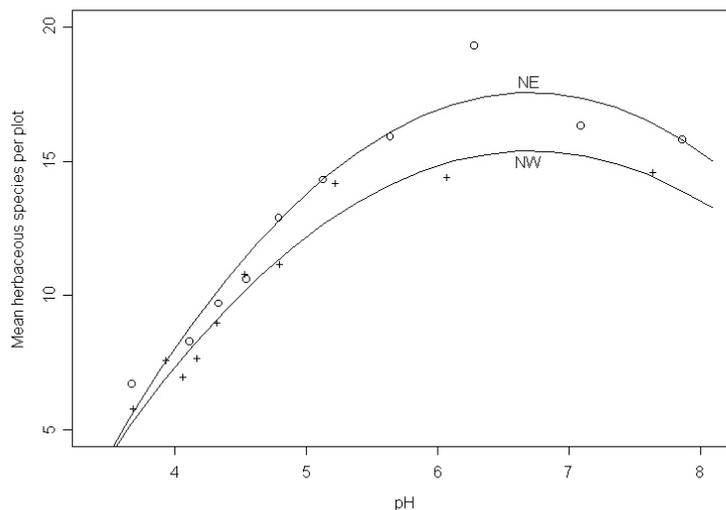
	Radical	OPT NW	CI95 NW	OPT NE	CI95 NE	AMP NW	AMP NE
<i>Convallaria majalis</i>	conma	6.0	[5.8, 6.5]	8.1	-	3.0	3.5
<i>Dryopteris carthusiana</i>	drca	5.1	[3.5, 5.6]	3.4	-	3.2	2.4
<i>Galeopsis tetrahit</i>	gate	6.2	[5.8, 7.3]	5.6	[5.4, 5.8]	2.9	3.5
<i>Oxalis acetosella</i>	oxac	5.0	[3.5, 5.6]	4.5	[3.9, 4.8]	2.9	2.1
<i>Teucrium scorodonia</i>	tesc	4.6	[4.3, 4.8]	3.4	-	2.1	2.7

**Changement pour OPT seul**

	Radical	OPT NW	CI95 NW	OPT NE	CI95 NE	AMP NW	AMP NE
<i>Athyrium filix-femina</i>	atfi	5.7	[5.4, 6.2]	5.1	[4.8, 5.3]	3.0	2.7
<i>Carex flacca</i>	cafla	7.1	[6.7, 8.1]	8.1	-	2.4	2.0
<i>Geranium robertianum</i>	gero	6.9	[6.6, 7.7]	6.3	[6.1, 6.7]	2.5	2.5
<i>Melica uniflora</i>	meun	6.6	[6.2, 7.8]	6.1	[5.9, 6.3]	2.9	2.8
<i>Mercurialis perennis</i>	mepe	8.1	-	7.5	[7.1, 8.1]	2.0	1.8
<i>Stachys sylvatica</i>	stsy	6.4	[6.1, 6.9]	8.1	-	2.4	2.3

**Changement pour AMP seul**

	Radical	OPT NW	CI95 NW	OPT NE	CI95 NE	AMP NW	AMP NE
<i>Carex remota</i>	care	5.9	[5.6, 6.5]	5.8	[5.6, 6.2]	2.9	2.4
<i>Luzula pilosa</i>	lupi	5.5	[5.2, 5.8]	5.4	[5.3, 5.5]	2.7	2.2
<i>Melampyrum pratense</i>	mepr	3.4	-	3.4	-	2.6	1.5
<i>Scrophularia nodosa</i>	scno	6.2	[6.0, 6.6]	6.0	[5.6, 6.5]	2.3	2.9



**Figure 4.25.** Nombres moyens d'espèces herbacées dans le NW et le NE le long du gradient de pH. Les relevés ont été agrégés en 10 groupes dans chaque région.

En terme d'amplitude écologique, seules les espèces acido-tolérantes ont révélé une claire tendance vers des valeurs plus élevées dans le NW par rapport au NE (par exemple *Melampyrum pratense*, *Oxalis acetosella* ou *Dryopteris carthusiana*), ceci car ces espèces peuvent s'adapter à une large gamme de milieux, depuis des milieux pauvres dans le NE ou le NW jusqu'à des milieux riches dans le NW où nous supposons que les conditions de compétitions sont moins contraignantes. Les pools régionaux d'espèces neutrophiles et les conditions de compétition (en termes de densité d'espèces) peuvent donc expliquer quelques observations faites sur la figure 4.23.

#### 4.4. Discussion et conclusions

Dans le nord de la France, la faible proportion observée d'espèces acidiphiles par rapport aux espèces neutrophiles, couplée à la faible proportions de sites forestiers basiques par rapport aux sites acides, illustre le "paradoxe d'Ewald" ("calcareous riddle", Ewald 2003) décrit pour l'Europe centrale : le fait d'observer de nombreux sites forestiers acides avec relativement peu d'espèces adaptées et peu de sites forestiers neutres à riches avec relativement beaucoup d'espèces adaptées.

Notre résultat principal, la stabilité relative de la niche par rapport au pH du sol pour la plupart des espèces forestières herbacées dans le nord de la France, est cohérent avec la plupart des études précédentes (par exemple Thompson et al. 1993, Hill et al. 2000, Prinzing et al. 2002). Néanmoins, nos résultats s'appuient sur des mesures réelles de pH et ont été obtenus grâce à deux jeux de données régionaux de grande qualité. Même si l'utilisation de valeurs indicatrices empiriques, telles les valeurs d'Ellenberg, ont pu mener à des résultats cohérents (par exemple Diekmann 1995, Ertsen et al. 1998, Dzwonko 2001), nous partageons l'idée que le fait de lier le comportement écologique des espèces forestières à des facteurs écologique mesurés, permettra d'augmenter notre connaissance de l'autécologie des espèces (Schaffers et Sykora 2000, Wamelink et al. 2002). Diekmann et Lawesson (1999) ont caractérisé le comportement écologique de huit espèces dans quatre régions distinctes, le long d'un gradient allant de l'Europe centrale au nord de l'Europe (gradient européen latitudinal) et ils ont montré que certaines espèces pouvaient révéler des changements d'optimum écologique dans différentes régions. Leur étude a été élargie par d'autres auteurs qui ont exploré des gradients longitudinaux en Europe (Hill et al. 1999, Prinzing et al. 2002, Gégout et Krizova 2003). Une étape future pourrait consister en la comparaison de nos résultats avec la littérature existante (Ellenberg et al. 1992, Hill et al. 1999, Gégout et Krizova 2003), pour vérifier si nos valeurs indicatrices de pH dans le nord de la France restent cohérentes le long d'un large gradient longitudinal en Europe. Il n'est probablement pas raisonnable d'affecter à toutes les espèces une valeur indicatrice unique qui serait valable pour toute l'Europe tempérée, mais il serait intéressant d'étudier plus en détails ces aspects de changements géographiques du comportement écologique à l'échelle du continent européen, en effectuant des analyses complémentaires à l'aide de grosses bases de données.

Nous avons montré que 11 espèces sur les 46 étudiées présentaient un léger changement vers des milieux de nutrition minérale favorables dans le NW (tableau 4.5). Toutefois, un artéfact technique, déjà pointé par Mohler (1983), pourrait être responsable d'un changement artificiel de l'optimum ou de l'amplitude vis-à-vis du pH pour quelques espèces, et ajouter ainsi une certaine variabilité à la caractérisation de la niche : en effet, nous avons observé que certaines espèces qui présentent un optimum à une extrémité du gradient de pH dans une régions présentaient souvent un changement de l'optimum dans l'autre région (par exemple *Dryopteris carthusiana*, *Teucrium scorodonia* à l'extrême inférieur du gradient de pH et *Mercurialis perennis*, *Convallaria majalis*, *Stachys sylvatica* et *Carex flacca* à l'extrême supérieur du gradient de pH). Les résultats révèlent que le changement d'optimum de ces espèces est cohérent, pour cinq d'entre elles sur les six, avec l'hypothèse d'un changement vers des conditions de ressource favorable dans le NW. Cet artéfact

peut probablement affecter l'ampleur du changement, mais il est peu probable qu'il biaise les principaux résultats.

Des modèles ajustés pour les mêmes espèces, mais dans des régions différentes, sont difficiles à comparer (Guisan et al. 2002) puisqu'ils modélisent simplement la niche réalisée et prennent en compte de manière implicite des contextes abiotiques et des interactions biotiques différents, qui sont susceptibles de varier entre les régions (Guisan et Zimmermann 2000). Dans notre étude, cependant, nous avons tenté de neutraliser l'effet des facteurs écologiques autres que le pH du sol (nutrition minérale), pour éviter d'examiner des changements de comportement écologique dans des contextes abiotiques trop disparates. Le fait que la relation entre le pH du sol et les autres facteurs de nutrition tels que le taux de saturation ou le rapport C/N soit stable entre le NW et le NE, nous permet de rejeter l'hypothèse qu'un facteur que nous n'avions pas pris en compte soit responsable des quelques changements observés pour certaines espèces.

L'hypothèse d'une différenciation d'écotypes pour la plupart des espèces entre le NW et le NE n'était pas réaliste, car elle ne peut pas être responsable de changements systématiques observés dans nos résultats. Les quelques changements observés ont ainsi été expliqués par des contextes biotiques différents dans les deux régions. Cette tendance à observer des optima/amplitudes de pH plus élevés dans le NW est probablement due à une libération de niche due à la compétition dans les milieux neutres à riches du NW, et nous suggérons que les espèces végétales sont soumises à des conditions de compétition plus fortes sur les sites forestiers du NE de pH compris entre 6 et 7. En moyenne, il y a trois espèces de plus par relevé dans le NE que dans le NW sur les sols neutres à basiques (figure 4.25), et nous avons aussi observé que le couvert herbacé total dans le NE était supérieur à 80 % pour des valeurs de pH supérieures à 6, et que ce couvert était plus important que le couvert herbacé total des forêts du NW, d'au moins 5 %, ceci tout le long du gradient de pH. Une explication de cette observation pourrait être une densité supérieures des forêts dans le NE par rapport au NW et donc une mobilité des plantes et une colonisation des milieux favorisées dans le NE (figure 4.21).

Notre étude traitait de la totalité des espèces herbacées fréquentes des forêts du nord de la France, et les résultats que nous avons générés n'auraient pu être observés en n'étudiant que quelques espèces. L'influence des conditions de compétition sur la réponse écologique des espèces a rarement été montrée en se basant sur des données réelles, et nous l'avons fait de manière indirecte dans cette étude : des études complémentaires peuvent alors être envisagées pour tester plus en détail les théories écologiques.



**5. Comparaison des pools d'espèces, de la richesse spécifique et de la réponse écologique des espèces forestières herbacées vis-à-vis du pH du sol dans les Vosges et le Jura**

## Résumé

L'objectif de cette étude était de caractériser les pools d'espèces forestières herbacées et la richesse spécifique, le long du gradient de pH, en se basant sur 745 relevés floristiques réalisés dans les Vosges et 205 relevés réalisés dans le Jura. Le comportement écologique des espèces présentes plus de 10 fois dans chaque massif a été caractérisé par des modèles de régression logistique simple et comparé en tenant compte des pools d'espèces présents dans chacun des massifs.

L'analyse des pools d'espèces herbacées dans les Vosges et le Jura a révélé une taille supérieure du pool d'espèces de milieux acides dans les Vosges et une taille supérieure du pool d'espèces de milieux riches dans le Jura, phénomène lié en premier lieu aux substrats dominants différents de ces deux massifs. La traduction des pools et de la richesse spécifiques en termes de compétition a révélé que les milieux neutres à riches (pH supérieur à 6) étaient les milieux subissant le plus de compétition.

Une grande majorité des espèces présentaient une réaction significative au facteur pH dans les Vosges (75 %) tandis qu'une minorité d'espèces dans le Jura (40 %) présentaient une réaction significative, phénomène probablement dû en partie au faible nombre de relevés disponibles dans le Jura. Parmi les 21 espèces présentes plus de 10 fois dans les deux massifs, une dizaine d'espèces présentaient un optimum de pH compris entre 6 et 7 dans les Vosges alors que l'optimum est compris entre 5 et 6 dans le Jura. Cette concentration d'espèces, pour les milieux présentant les conditions nutritionnelles les plus favorables dans les Vosges, est probablement liée à une compétition moins forte pour ces milieux que dans le Jura.

## *Mots clés*

Jura, Vosges, espèces forestières herbacées, pool, richesse, régions, pH, régression logistique, compétition.

## 5.1. Introduction

La grande diversité des conditions écologiques rencontrées en milieu de montagne a souvent été décrite. A cette diversité écologique s'ajoute une grande diversité floristique qui lui est en partie liée (Körner 1995). L'étude simultanée de l'écologie des espèces dans plusieurs massifs de montagne a été effectuée principalement par des analyses multivariées (Allen et al. 1991, Bergmeier 2002) et les variables écologiques étudiées depuis les années 1990 dans un contexte de montagne, sont des variables climatiques issues de croisements sous systèmes d'information géographique (SIG). En effet, dans un contexte de changement global, c'est souvent la réponse à un facteur de température qui est étudiée, pour prédire notamment la réponse des espèces et des communautés végétales à une augmentation de ce facteur (Brzeziecki et al. 1994, Bugmann 1994, Körner 1995, Guisan et al. 1998, Guisan et Theurillat 2000a, 2000b).

Peu d'études concernent une caractérisation édaphique du comportement des espèces en montagne. De plus, dans les montagnes françaises, le comportement nutritionnel des espèces a été étudié de façon inégale selon les massifs. Le massif vosgien a fait l'objet de plusieurs études (par exemple Penel 1979, Gégout 1995), qui montrent que le gradient trophique représente le facteur majeur de discrimination des espèces et des communautés végétales dans la région. Le niveau trophique des milieux, souvent synthétisé par le facteur pH eau du sol, représente donc avec l'altitude, un facteur déterminant de discrimination des espèces et des communautés végétales dans les forêts de montagne.

En France, il n'existe pas d'études s'étant attaché à examiner le comportement nutritionnel des espèces dans différents massifs de montagne et la question de la comparaison de la réponse écologique des espèces aux facteurs du sol n'a pas été traitée, bien que de nombreuses conditions biotiques et abiotiques soient différentes selon les massifs. En effet, la composition floristique et les relations de compétition varient selon les massifs, et l'isolement des massifs les uns par rapport aux autres entraîne un isolement des populations d'espèces montagnardes qui peut a priori entraîner des variations de comportement écologique des espèces. Une analyse des différences et des ressemblances de comportement écologique des espèces rencontrées dans différentes montagnes françaises permettrait de tester cette hypothèse et d'avoir une vision d'ensemble du comportement trophique des espèces à l'échelle de plusieurs massifs de montagne.

L'objectif de ce chapitre est de caractériser, dans un premier temps, les pools d'espèces et la richesse spécifique des espèces forestières herbacées, le long du gradient pH, dans les Vosges et le Jura afin d'appréhender une composante importante de la compétition. Dans un second temps, l'objectif est de caractériser, par une méthode de régression paramétrique adaptée, la régression logistique, la réponse des espèces forestières vis-à-vis du pH dans ces deux massifs de montagne du nord-est de la France. En particulier, l'hypothèse testée, liée à un changement attendu de comportement des espèces face au facteur pH, est la suivante : des phénomènes de compétition différents selon les massifs conduisent à un décalage des optima des espèces vers des milieux plus favorables lorsque la compétition est réduite (voir le chapitre précédent).

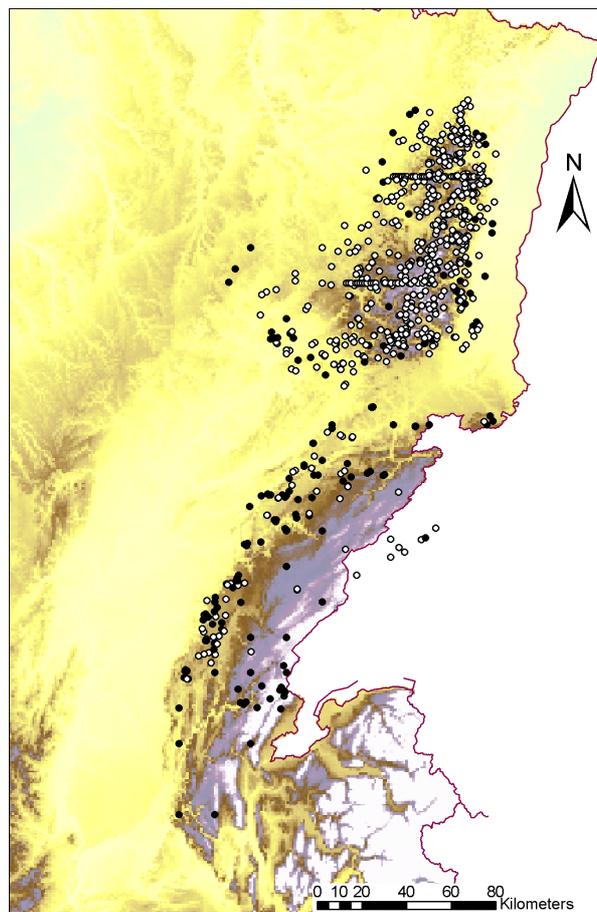
La procédure d'analyse du comportement écologique des espèces par rapport au pH dans les Vosges et le Jura, sera la suivante : dans un premier temps, nous examinerons les conditions de compétition au sein de chaque massif, par l'intermédiaire de l'analyse des pools d'espèces et des variations de richesse spécifique le long du gradient pH. La compétition sera considérée comme d'autant plus importante que le pool et la richesse spécifiques seront grands. Dans un second temps, nous caractériserons la réponse des espèces par rapport au pH dans chaque massif, grâce à des courbes de réponse issues de modèles de régression logistique gaussienne, puis nous extrairons de ces courbes l'optimum et l'amplitude écologiques. Enfin, dans un dernier temps,

nous comparerons la réponse des espèces présentes dans les deux massifs pour tester l'hypothèse énoncée.

## 5.2. Matériels et méthodes

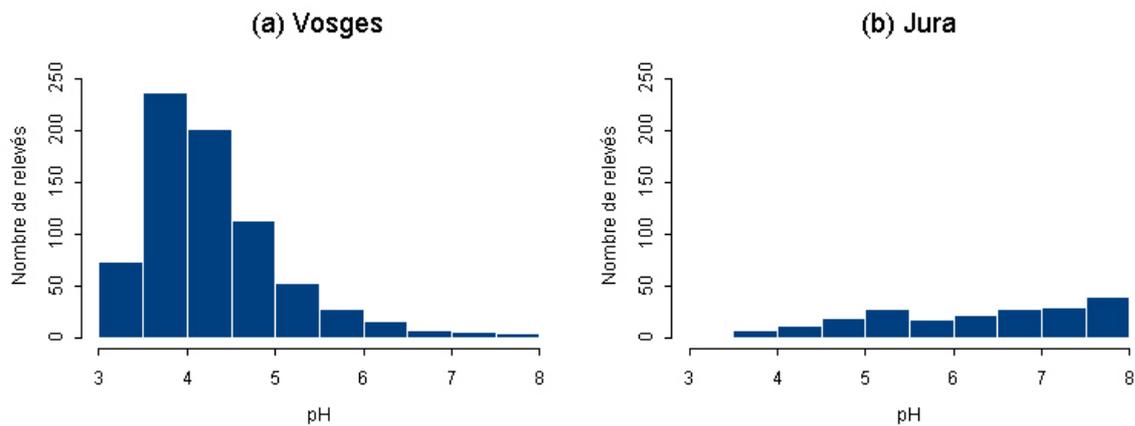
### 5.2.1. Régions d'étude et données disponibles

Deux régions de montagnes françaises font l'objet de cette étude, les Vosges et le Jura, situées dans le nord-est de la France (figure 5.26). Les relevés phytocécologiques réalisés dans ces massifs ont été extraits de la base EcoPlant (Gégout 2001, Gégout et al. 2005), afin de bénéficier respectivement de 745 et 205 relevés dans les Vosges et le Jura, issus respectivement de 22 et 12 études phytocécologiques.



**Figure 5.26.** Distribution spatiale des placettes comportant un relevé floristique et une mesure de pH du premier horizon du sol dans les Vosges (745 relevés) et le Jura (205 relevés). Les points blancs et noirs correspondent aux placettes pour lesquelles le pH eau de l'horizon A est respectivement inférieur et supérieur à 5,5.

La gamme de pH observée dans les relevés des Vosges est [3,1 - 7,8] et elle est de [3,6 - 8,0] pour les relevés du Jura (figure 5.27).



**Figure 5.27.** Distribution des pH des relevés disponibles dans (a) les Vosges (745 relevés) et (b) le Jura (205 relevés).

C'est dans le massif vosgien, à forte dominante acide, que le nombre de relevés est le plus important (745), probablement dû à la proximité géographique des organismes d'enseignement et de recherche présents en Lorraine. L'altitude des relevés s'étend de 230 à 1320 mètres dans le jeu de données des Vosges et de 212 à 1420 mètres dans celui du Jura.

## 5.2.2. Caractérisation de la composition floristique

### 5.2.2.1. Pools régionaux

Le pool d'espèces joue un rôle important dans la régulation de la richesse spécifique des relevés (Zobel 1997, Grace 1999, Pärtel 2002). Zobel et al. (1998) définissent le pool régional d'espèces comme étant le réservoir de toutes les espèces physiologiquement capables de vivre dans les conditions écologiques d'une certaine communauté. Cependant, il est très difficile de caractériser les pools d'espèces de manière statistique (Eriksson 1993, Pärtel et al. 1996, Dupré 2000, Grace 2001).

Dans chacun des deux massifs, et en se basant sur les jeux de données disponibles, on peut définir le pool d'espèces comme un réservoir d'espèces potentiellement présentes. Dans les 745 relevés des Vosges et les 205 relevés du Jura, on trouve respectivement 308 et 363 espèces herbacées, qui représentent les tailles totales des pools d'espèces herbacées. La plupart des espèces sont des espèces peu fréquentes puisque le nombre d'espèces présentes dans plus de 10 relevés est de 102 dans les Vosges, pour 92 dans le Jura ; si l'on considère les espèces présentes dans plus de 50 relevés, le pool d'espèces est alors de 42 espèces dans les Vosges et 20 dans le Jura. Cependant, ces pools d'espèces sont difficilement comparables directement car ils ont été établis avec des nombres de relevés très différents dans chaque massif.

Pour évaluer la taille d'un pool d'espèces en fonction du nombre de relevés floristiques, les relevés ont été regroupés dans un tableau floristique à  $n$  lignes (relevés) et  $p$  colonnes (espèces). Parmi tous les relevés, le principe de la méthode a consisté à tirer au hasard  $r$  relevés, puis à compter le nombre d'espèces différentes dans ces  $r$  relevés. Par exemple, on peut tirer au hasard 10 relevés ( $r = 10$ ) parmi les 745 disponibles dans les Vosges et compter combien d'espèces différentes sont contenues dans ces 10 relevés. Pour connaître l'influence du nombre de relevés tirés au hasard sur le nombre d'espèces présentes, on peut faire varier la valeur  $r$  entre 1 et le nombre de relevés disponibles. Lorsque  $r$  vaut 1, cela revient à compter le nombre d'espèces présentes dans un relevé (c'est la richesse spécifique du relevé) et lorsque  $r$  vaut le nombre de

relevés disponibles (745 pour les Vosges et 205 pour le Jura), le nombre d'espèces constitue le pool total des espèces.

Nous avons donc fait varier le nombre de relevés tirés au hasard entre 1 et 745 pour les Vosges et entre 1 et 205 pour le Jura, de cinq en cinq relevés (1, puis 5, puis 10, puis 15, etc.). A chaque valeur de  $r$ , nous avons effectué 10 itérations du calcul et calculé le nombre d'espèces présentes dans les  $r$  relevés pour chacune de ces 10 itérations, puis calculé la moyenne. Une courbe résultante permet d'observer le nombre moyen d'espèces présentes (la taille moyenne du pool) en fonction du nombre de relevés tirés au hasard.

#### 5.2.2.2. Pools régionaux et pH

La détermination de la taille du pool d'espèces a été présentée ci-dessus sans tenir compte d'aucun gradient écologique. Cependant, peu d'études ont caractérisé le pool d'espèces disponibles le long d'un gradient écologique, plus particulièrement le long du gradient de pH (Peet et al. 2003). En Europe centrale, Ewald (2003) a montré que le pool d'espèces vasculaires était beaucoup plus grand sur sols riches et calcaires que sur sols acides, en se basant sur la préférence écologique des espèces exprimées par la valeur R d'Ellenberg et al. (1992).

Nous avons donc déterminé le nombre moyen d'espèces rencontrées dans 10 relevés lorsque l'on se déplace le long du gradient de pH. Pour cela, dans chaque unité de pH, nous avons tiré au sort 10 relevés et déterminé le nombre d'espèces présentes dans ces 10 relevés. L'opération a été renouvelée 100 fois pour chaque unité de pH pour calculer un pool moyen d'espèces dans 10 relevés de chaque région.

#### 5.2.2.3. Similarité des relevés floristiques et pH

Il existe de nombreux indices pour caractériser la (dis)similarité de deux relevés floristiques (Gower et Legendre 1986). Pour comparer la composition floristique des deux massifs en tenant compte à la fois des espèces présentes et des espèces absentes de chaque massif, nous avons utilisé l'indice de similarité de Dice, dont la définition est rappelée ci-dessous (équation 5.8) :

$$2 \cdot n_c / (n_1 + n_2) \quad \text{(Equation 5.8)}$$

où  $n_c$  est le nombre d'espèces communes à deux relevés floristiques  $r_1$  et  $r_2$  et  $n_1$  et  $n_2$  les nombres d'espèces présentes respectivement dans les relevés  $r_1$  et  $r_2$ .

Pour évaluer la similarité des relevés floristiques le long du gradient de pH dans les Vosges et le Jura, les relevés ont été classés par ordre croissant de pH et groupés par effectifs de 20 relevés. L'indice de similarité de Dice peut alors être calculé pour toutes les combinaisons de paires possibles entre ces 20 relevés et moyenné pour obtenir une valeur de l'indice que l'on peut associer au pH moyen des 20 relevés.

Ainsi, pour les Vosges (745 relevés), 37 doublets (indice de Dice moyen, pH moyen) ont pu être calculés, contre 10 pour le Jura (205 relevés).

Pour caractériser la (dis)similarité de relevés floristiques entre les Vosges et le Jura sur des pH comparables, nous avons tiré au hasard 10 relevés floristiques des Vosges et 10 relevés floristiques du Jura dans toutes les gammes d'une unité de pH ([3,0 - 4,0], [3,5 - 4,5], [4,0 - 5,0], ..., [6,5 - 7,5], [7,0 - 8,0]), puis calculé l'indice de similarité de Dice pour toutes les combinaisons de paires possibles par classe de pH. En associant l'indice moyen de toutes les combinaisons

possibles au pH moyen des 20 relevés sélectionnés, neuf doublets (indice Dice moyen, pH moyen) ont pu être calculés.

#### 5.2.2.4. Richesse spécifique et pH

Le nombre d'espèces présentes dans un relevé dépend tout naturellement en premier lieu de la taille du relevé, mais les jeux de données réunis pour les Vosges et le Jura présentent des tailles de relevés similaires (400 m<sup>2</sup>), ce qui permet d'estimer la richesse spécifique sans biais apparent.

Les différences de richesse spécifique le long d'un gradient écologique particulier n'ont pas été beaucoup étudiées (voir cependant Dupré et al. 2002, Schuster et Diekmann 2003, 2005). Pärtel (2002) a montré qu'en général, la richesse spécifique était positivement liée à la valeur du pH dans les zones tempérées, et Schuster et Diekmann (2003) ont montré que la relation entre le pH et la richesse spécifique du relevé était croissante ou unimodale, avec un pic de richesse pour des pH proches de 5 en forêt en Allemagne.

Pour caractériser la richesse spécifique herbacée le long du gradient pH, nous avons calculé le nombre d'espèces herbacées présentes dans chaque relevé, puis avons ajusté un modèle de régression quadratique entre le nombre d'espèces dans le relevé et le pH correspondant (Schuster et Diekmann 2003).

### 5.2.3. Caractérisation du comportement écologique

Pour chaque espèce présente plus de 10 fois dans un massif, on peut ajuster une courbe de réponse avec la technique de régression logistique simple (cf. chapitres précédents). De cette courbe, on tire des résumés numériques : l'optimum écologique, l'amplitude écologique et la probabilité maximale de la courbe.

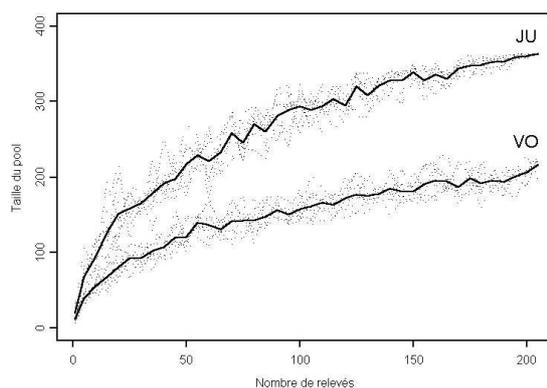
## 5.3. Résultats

### 5.3.1. Composition floristique dans les Vosges et le Jura

#### 5.3.1.1. Pools régionaux

La mesure de richesse floristique la plus simple est la taille du pool d'espèces présentes dans chaque massif. Dans les Vosges (745 relevés) et le Jura (205 relevés), respectivement 308 et 363 espèces herbacées ont pu être relevées. Avec environ trois fois moins de relevés floristiques disponibles dans le Jura, on note que le pool d'espèces disponibles est toutefois supérieur dans ce massif par rapport à celui des Vosges.

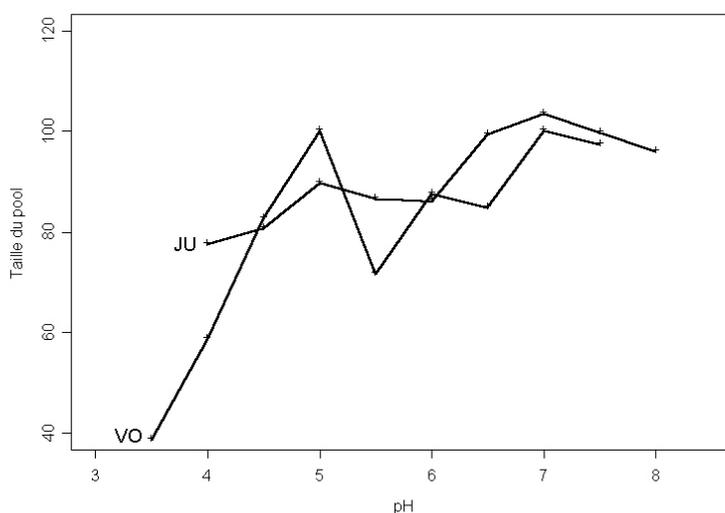
La taille du pool d'espèces ne varie pas de la même manière dans les deux massifs, selon le nombre de relevés floristiques considérés (figure 5.28). La différence de taille de pool est plus nette lorsque l'on considère le même nombre de relevés : dans les 205 relevés floristiques du Jura, on compte 363 espèces différentes, tandis que dans 205 relevés floristiques des Vosges, on compte en moyenne 217 espèces différentes (calcul basé sur la moyenne de 10 tirages successifs avec remise de 205 relevés floristiques parmi les 745 relevés disponibles, voir le chapitre matériels et méthodes).



**Figure 5.28.** Taille du pool d'espèces herbacées dans les Vosges et le Jura, en fonction du nombre de relevés floristiques. Les courbes en gras représentent les courbes moyennes de 10 estimations (en pointillés), issues chacune de tirages successifs du nombre de relevés considéré parmi tous les relevés disponibles (745 pour les Vosges et 205 pour le Jura).

L'analyse des pools régionaux laisse apparaître une plus grande richesse spécifique herbacée dans le Jura par rapport au massif vosgien. Il est nécessaire d'examiner si ce phénomène n'est pas lié à la distribution très différente des pH dans les Vosges et dans le Jura.

#### 5.3.1.2. Pools régionaux et pH

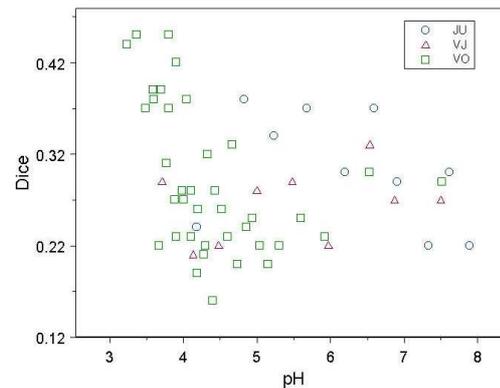


**Figure 5.29.** Taille du pool des espèces herbacées dans les Vosges (VO) et le Jura (JU), le long du gradient pH.

Une fois tenu compte du pH, le pool d'espèces semble globalement comparable d'une région à l'autre (un peu plus élevé dans les Vosges pour les pH de 4 à 5 et un peu plus élevé dans le Jura pour les pH de 6 à 7). Au delà d'un pH 4,5, le pool oscille entre 80 et 100 espèces pour 10 relevés dans les deux régions. Un changement de comportement des espèces entre les deux massifs ne pourrait donc pas être attribué à une différence de pools d'espèces.

### 5.3.1.3. Similarité des relevés floristiques et pH

Pour les basses valeurs de pH, des valeurs relativement importantes sont atteintes pour l'indice de similarité des relevés floristique des Vosges, ce qui exprime une certaine redondance floristique. Pour des valeurs intermédiaires à fortes de pH, ce sont les relevés floristiques du Jura qui semblent les plus similaires (figure 5.30).

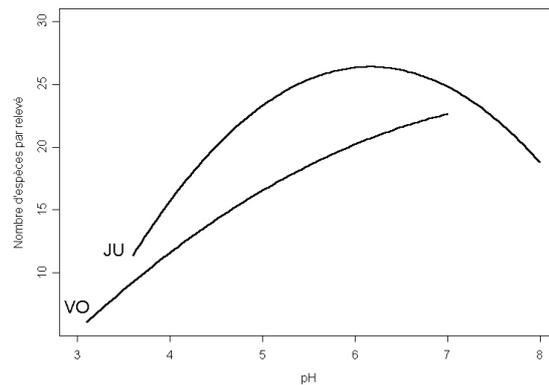


**Figure 5.30.** Indices moyens de similarité de Dice des relevés floristiques le long du gradient de pH, pour les relevés des Vosges (VO), du Jura (JU) et de l'ensemble des deux massifs (VJ). La procédure de calcul de l'indice est présentée dans le chapitre Matériels et méthodes.

La similarité est la même au sein des régions qu'entre les régions. On n'attend donc pas de changement de compétition lié à un changement de composition floristique d'une région à l'autre.

### 5.3.1.4. Richesse spécifique et pH

La richesse spécifique par relevé, pour les deux massifs, est minimale sur les pH les plus acides et la relation avec le pH est croissante dans les Vosges et en cloche dans le Jura (figure 5.31). Dans toute la gamme de pH, le nombre d'espèces par relevé est supérieur dans le Jura par rapport aux Vosges et l'écart est maximal pour les pH intermédiaires situés dans la gamme [5,0 - 6,5]. On peut donc penser que les conditions de compétition sont plus fortes dans le Jura que les Vosges.



**Figure 5.31.** Nombre d'espèces forestières herbacées par relevé le long du gradient pH (VO, Vosges et JU, Jura). La courbe est issue d'une régression quadratique entre les données brutes de pH et de richesse spécifique associée, pour les 745 relevés des Vosges et les 205 relevés du Jura.

#### 5.3.1.5. Bilan de l'étude de la composition spécifique herbacée

Les différents résultats obtenus, en termes de composition spécifique dans les Vosges et le Jura ont révélé que :

- (i) la taille globale du pool d'espèces est plus grande dans le Jura que dans les Vosges ;
- (ii) la taille du pool d'espèces est comparable entre les deux régions une fois tenu compte du pH ;
- (iii) la composition floristique à pH fixé ne varie pas plus entre les deux massifs qu'en leur sein ;
- (iv) en règle générale, il y a plus d'espèces par relevé dans le Jura par rapport aux Vosges, quelle que soit la position le long du gradient de pH.

Le pool d'espèces des Vosges semble donc inférieur à celui du Jura globalement mais ceci n'est plus vrai lorsque l'on tient compte du pH. Par contre, la richesse spécifique des relevés du Jura est supérieure à celle des relevés des Vosges spécialement dans les milieux riches. La compétition pour les milieux à bonne nutrition minérale (pH 6 - 7) est supérieure dans le Jura par rapport aux Vosges.

### 5.3.2. Réponse des espèces au pH

#### 5.3.2.1. Courbes de réponse

Dans les Vosges, 102 espèces forestières herbacées étaient présentes plus de 10 fois parmi les 745 relevés disponibles, et un modèle de régression logistique a pu être défini pour toutes ces espèces. Parmi ces 102 espèces, 25 ne réagissent pas au facteur pH (25 %). Dans le Jura, 92 espèces forestières herbacées étaient présentes parmi les 205 relevés disponibles et 55 espèces parmi les 92 (60 %) ne présentent pas de réaction par rapport au facteur pH. 77 courbes de réponse pour les Vosges et 37 courbes pour le Jura ont pu être ainsi dérivées. 21 espèces présentaient une réaction significative au pH dans les deux massifs (figure 5.32). La distribution des optima des espèces qui présentaient une réaction significative au pH dans chaque région révèle un pic d'espèces avec un optimum pour des pH intermédiaires dans les Vosges (pH entre 5 et 7), ainsi qu'un pic d'espèces avec un optimum à pH maximum dans le Jura (figure 5.33).

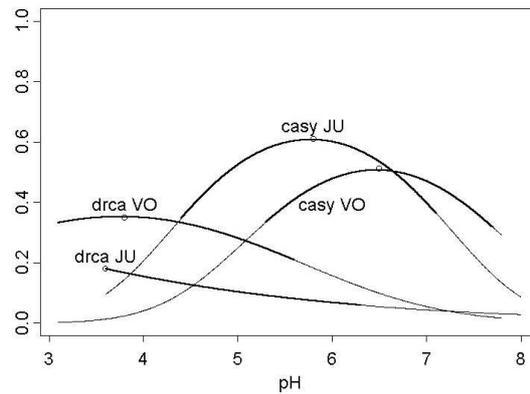


Figure 5.32. Exemples de courbes de réponse par rapport au pH, de *Dryopteris carthusiana* (drca) et de *Carex sylvatica* (casy), dans les Vosges (VO) et le Jura (JU). L'optimum de l'espèce est marqué par un point et l'amplitude écologique correspond au trait épais.

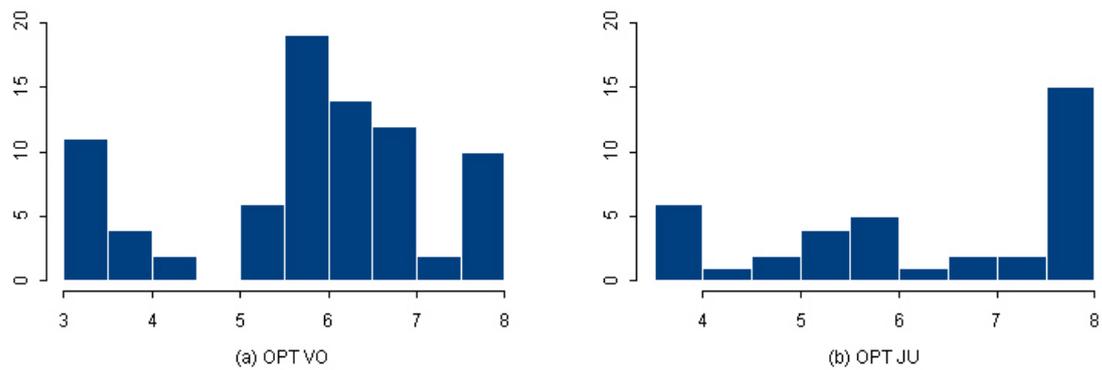


Figure 5.33. Histogramme des optima des espèces présentes plus de 10 fois et présentant une réaction significative au pH dans les Vosges (VO, 77 espèces) et le Jura (JU, 37 espèces).

### 5.3.2.2. Comparaison des optima de pH dans les Vosges et le Jura

Un optimum de pH a pu être déterminé dans les deux massifs pour 21 espèces présentes plus de 10 fois dans chaque massif (figure 5.34).

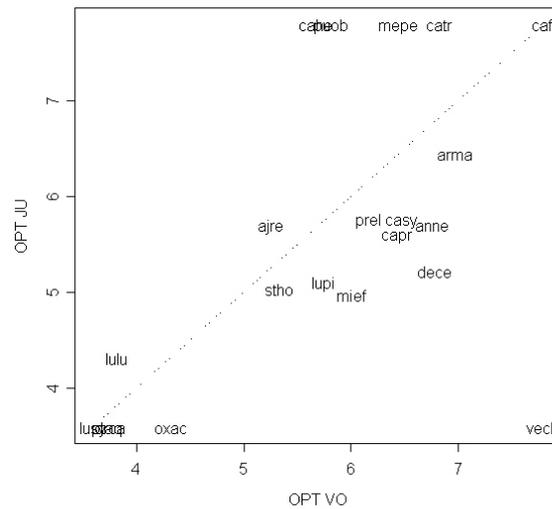


Figure 5.34. Optimum de pH dans les Vosges (OPT VO) et le Jura (OPT JU) pour 21 espèces forestières herbacées présentant une réaction significative au facteur dans les deux massifs. Les espèces sont reconnues par un code défini dans le texte.

Cinq groupes d'espèces peuvent être définis :

- (i) cinq espèces acidiphiles dans les deux massifs : *Luzula sylvatica* (lusu), *Pteridium aquilinum* (ptaq), *Dryopteris carthusiana* (drca), *Oxalis acetosella* (oxac) et *Luzula luzuloides* (lulu) ;
- (ii) une espèce calcaricole dans les deux massifs : *Carex flacca* (cafl) ;
- (iii) une espèce présentant deux optima totalement opposés dans les deux massifs : *Veronica chamaedrys* (vech) ; cependant la faible fréquence de cette espèce dans les deux massifs impose une certaine réserve quant à la validité du résultat (11 occurrences dans les Vosges et 14 occurrences dans le Jura) ;
- (iv) 10 espèces présentant une réaction intermédiaire dans les deux massifs : *Milium effusum* (mief), *Stellaria holostea* (stho), *Luzula pilosa* (lupi), *Deschampsia cespitosa* (dece), *Cardamine pratensis* (capr), *Anemone nemorosa* (anne), *Ajuga reptans* (ajre), *Primula elatior* (prel), *Carex sylvatica* (casy) et *Arum maculatum* (arma) ; dans ce groupe on note la concentration d'espèces avec un optimum de pH entre 6 et 7 dans les Vosges et avec un optimum entre 5 et 6 dans le Jura.
- (v) quatre espèces avec un optimum intermédiaire dans les Vosges et un optimum extrême supérieur dans le Jura : *Cardamine heptaphylla* (cahe), *Pulmonaria obscura* (puob), *Mercurialis perennis* (mepe) et *Campanula trachelium* (catr) ; ces quatre espèces présentent un optimum de pH dans les Vosges compris entre 6 et 7.

#### 5.4. Discussion

Malgré un pool d'espèces légèrement plus important dans les milieux acides, les Vosges ont moins d'espèces par relevé sur ces milieux, signifiant que la taille du pool régional n'est pas forcément corrélée de manière positive avec la richesse spécifique à l'échelle du relevé (Ewald 2003). En effet, la taille du pool d'espèces confinées dans des milieux acides (pH < 5,5) était de 90 espèces dans les Vosges et de 27 espèces dans le Jura, mais le nombre d'espèces par relevé était supérieur dans le Jura dans ces mêmes milieux. Dans les milieux plus riches, la taille du pool

légèrement plus importante dans le Jura que dans les Vosges illustre le fait que le nombre d'espèces calcicoles est plus important que le nombre d'espèces acidiphiles dans les forêts d'Europe, c'est le "calcareous riddle" décrit par [Ewald \(2003\)](#).

La forme unimodale de la relation entre nombre d'espèces par relevé et pH peut être due à l'effet d'autres facteurs écologiques et de productivité, mais aussi au grand recouvrement des amplitudes des espèces au milieu du gradient d'acidité ([Schuster et Diekmann 2003](#)). La réponse physiologique des plantes à l'acidité peut également expliquer la forme de courbe en cloche de la richesse spécifique par rapport au pH, puisque pour des pH trop élevés, trop peu de fer, de manganèse et de phosphate sont disponibles pour la plante et pour des pH trop bas, une trop grande concentration en ions hydrogène et une toxicité aluminique peuvent empêcher la survie de beaucoup d'espèces ([Tyler 1999](#)).

Plus des trois quarts des espèces étudiées présentent une réaction significative au pH dans les Vosges. Une proportion relativement faible d'espèces présentait une réaction significative au gradient de pH dans le Jura. Cela est probablement lié au faible nombre de relevés disponibles dans ce massif, comme l'a montré le chapitre étudiant la réponse des espèces en fonction du nombre d'occurrences prises en compte (chapitre 3). Cependant, cette variation du nombre d'occurrences prises en compte dans chaque massif n'explique pas l'ensemble des différences d'optima observées entre les Vosges et le Jura.

La compétition est souvent considérée comme l'un des principaux déterminants de la diversité des espèces ([Grime 1979](#)), mais elle a aussi une grande influence sur leur comportement écologique. Les conditions de nutrition minérale dans les forêts de France sont optimales pour des valeurs de pH autour de 6 - 7, comme on l'a observé dans le chapitre précédent. Une plus grande richesse spécifique par relevé est associée à ces milieux d'un bon niveau trophique, et les conditions de compétition y sont les plus contraignantes pour les plantes. On observe pour les autres espèces une tendance systématique de déplacement des optima vers des pH entre 6 et 7 dans les Vosges. Si on considère que la zone de pH située entre 6 et 7 est la plus favorable pour la nutrition minérale des plantes dans les forêts de France (voir chapitre précédent), les espèces adaptées à ces conditions vont avoir tendance à y trouver leur optimum. Dans les Vosges, le pool d'espèces et la richesse spécifique étant plus faibles que dans le Jura, les conditions de compétition sont plus faibles et les espèces peuvent avoir leur fréquence maximale (leur optimum) dans les milieux les plus favorables. Par contre, le pool et le nombre d'espèces importants dans le Jura pour les pH 6 - 7 empêchent la plupart d'entre elles d'avoir leur fréquence maximale (et donc leur optimum) pour ces valeurs de pH.

Un effet d'île, isolant chaque massif et ne favorisant pas l'échange d'individus d'une même population, aurait pu entraîner une plus grande variabilité de comportement chez les espèces inféodées aux montagnes, par rapport aux espèces ubiquistes géographiquement, que l'on retrouve également dans les plaines françaises. Cependant, la comparaison du comportement d'espèces montagnardes avec des espèces ubiquistes géographiquement n'a révélé aucune tendance particulière. Ce résultat suggère que les populations isolées ne semblent pas plus différer dans leur comportement écologique par rapport au pH que les populations ubiquistes. Des études ultérieures, menées sur des massifs plus éloignés les uns des autres et avec un plus grand nombre de relevés disponibles, devront être mises en œuvre pour vérifier cette hypothèse.



## **Troisième partie**

Etude de la niche écologique des espèces à l'échelle de la France



**6. Les facteurs nutritionnels du sol améliorent les modèles de prédiction de la distribution des espèces végétales : une illustration avec l'Erable champêtre (*Acer campestre* L.) en France**

*Article associé*

**Coudun, Ch., J.-C. Gégout, C. Piedallu et J.-C. Rameau (sous presse)** Soil nutritional factors improve plant species distribution models: an illustration with *Acer campestre* L. in France. *Journal of Biogeography*.

## Résumé

A travers la modélisation prédictive de la distribution de l'Erable champêtre (*Acer campestre* L.) dans les forêts de France, nous souhaitons illustrer le besoin de prendre en compte des variables édaphiques de nutrition lorsque l'on modélise la distribution d'espèces végétales.

Nous avons utilisé de l'information de présence/absence d'*Acer campestre* dans 3 286 relevés forestiers disséminés sur toute la France, couplée à de l'information climatique et édaphique. Plus de 150 variables climatiques (températures, précipitations, radiations solaires, évapotranspiration, bilan hydrique) ont été obtenues à l'aide d'un système d'information géographique (SIG), et six variables édaphiques directes (pH, C/N, taux de saturation S/T, concentrations en calcium, magnésium et potassium) étaient disponibles grâce à EcoPlant, une base de données phytoécologiques pour les forêts de France (Gégout et al. 2005). En utilisant la régression logistique pas-à-pas ascendante, nous avons calculé deux modèles prédictifs distincts de la distribution d'*Acer campestre*, le premier n'incluant que des variables climatiques, le second incluant des variables climatiques et des variables édaphiques.

Nous avons trouvé que la distribution d'*Acer campestre* était difficile à modéliser en n'incluant que des variables climatiques (taux de succès global de prédiction de la présence/absence,  $S = 55\%$ ), et que la prise en compte de variables édaphiques permettait d'augmenter significativement la qualité de la prédiction de l'espèce ( $S = 81\%$ ), en permettant de prédire des taches de présence/absence à des échelles locales. Nous concluons que des variables édaphiques de nutrition peuvent améliorer les modèles de prédiction de la distribution de certaines espèces.

## Mots-clés

*Acer campestre*, modélisation de la distribution des espèces, cartographie prédictive, régression logistique, autocorrélation spatiale, variables climatiques, variables de nutrition, EcoPlant, France.

## 6.1. Introduction

La modélisation prédictive des espèces et des habitats constitue aujourd'hui un domaine de l'écologie qui évolue très rapidement, puisque plusieurs centaines d'études ont été menées depuis la fin des années 1990s (Franklin 1995, Guisan et Zimmermann 2000, Guisan et al. 2002, Lehmann et al. 2002a, Scott et al. 2002, Rushton et al. 2004, Austin soumis). Plus particulièrement, les plantes terrestres ont souvent fait l'objet de modélisation de la distribution (voir par exemple Carpenter et al. 1993, Lenihan 1993, Guisan et al. 1998, 1999, Zimmermann et Kienast 1999, Hooten et al. 2003, Randin et al. sous presse). L'information sur la distribution des espèces et des habitats est nécessaire dans plusieurs optiques, telles que les questions d'invasion (Peterson 2003, Peterson et al. 2003), l'évaluation des conséquences des changements globaux (Iverson et al. 1999, Schwartz et al. 2001, Theurillat et Guisan 2001, Bakkenes et al. 2002, Dirnböck et al. 2003, Skov et Svenning 2004, Araújo et al. sous presse), le test d'hypothèses biogéographiques (Huntley et al. 2004, McPherson et al. 2004), ou les questions de conservation (Margules et Stein 1989, Araújo et Williams 2000, Polasky et al. 2000, Araújo et al. 2004, Cabeza et al. 2004, Williams et al. 2005). La combinaison de modèles de plantes particulières peut également conduire à la cartographie des communautés et des types de végétation (Cawsey et al. 2002, Elith et al. 2002).

La plupart des modèles actuels reposent sur l'hypothèse d'un pseudo-équilibre entre les espèces et leur environnement (Leathwick 1998, Guisan et Zimmermann 2000, Araújo et Pearson 2005), c'est-à-dire que tous les habitats qui sont favorables à une espèce sont colonisés par celle-ci. De nombreuses techniques statistiques ont été développées pour modéliser la distribution des espèces vis-à-vis de facteurs environnementaux (Yee et Mitchell 1991, Huisman et al. 1993, Elith 2000, Vayssières et al. 2000, Gelfand et al. 2003, Robertson et al. 2004). Parmi toutes les techniques, les modèles linéaires généralisés (GLMs, McCullagh et Nelder 1997) et les modèles additifs généralisés (GAMs, Hastie et Tibshirani 1997) sont les techniques les plus utilisées actuellement (Bio et al. 1998, Bio 2000, Austin 2002, Ferrier et al. 2002, Luoto et al. sous presse). Des logiciels spécifiques adaptés ont été développés (par exemple GRASP, Lehmann et al. 2002c).

La réponse écologique et la distribution spatiale des plantes sont souvent étudiées par rapport à des facteurs climatiques uniquement. Ces facteurs sont considérés très importants pour expliquer la distribution des espèces à de larges échelles (Box et al. 1993, Shao et Halpin 1995, Heegaard 2002b, Lehmann et al. 2002b). En effet, avec le développement des systèmes d'information géographique (SIG), l'accès à des couches de données climatiques interpolées à partir de mesures de stations météorologiques, est facilité (McKenzie et Halpern 1999, McKenzie et al. 2003a, 2003b, Thuiller et al. 2003b, 2004). Cependant, par définition, les facteurs du sol sont également très importants pour les espèces végétales, mais ils sont souvent négligés, par un manque de données précises disponibles. Par conséquent, l'étude simultanée de la réponse des plantes par rapport à la fois à des facteurs climatiques et édaphiques a rarement été menée (voir cependant Austin 1971, 1972, Gignac et al. 1991, Gignac 1992, Bragazza et Gerdol 1996, Heegaard 1997, 2001, Pinto et Gégout, 2005).

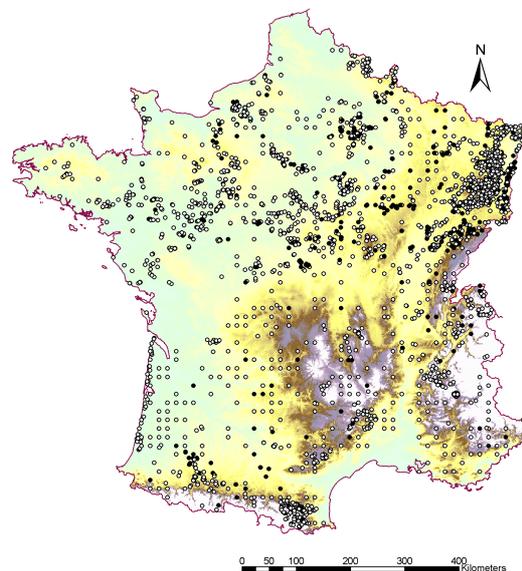
Dans ce chapitre, notre objectif principal était de tester l'apport des facteurs édaphiques en complément de facteurs climatiques pour expliquer la distribution d'une espèce d'arbre en France, en utilisant la régression logistique. Nous avons choisi l'Erable champêtre (*Acer campestre*) parce que cette espèce est connue pour être sensible à la fois à des facteurs du sol et du climat. *Acer campestre* L. (Erable champêtre) est un arbre de 12-15 m, que l'on peut rencontrer dans des bocages et des forêts mélangées à feuilles caduques dans presque toute l'Europe (Rameau et al.

1989, Mills 1996), mais qui est absent des zones nordique et méditerranéenne (Tutin et al. 2001). Même si ses exigences autécologiques précises restent mal connues, cette espèce présente une préférence pour des climats tempérés et des sols riches (souvent calcaires), et se trouve souvent sur des stations sèches (Motel 1995, Bendixen 2001). En France, on trouve *Acer campestre* partout, sauf dans la plupart des montagnes et dans les régions atlantique et méditerranéenne (Rameau et al. 1989). Depuis les années 1990, cet arbre a suscité un intérêt économique en France, mais il reste non géré par les forestiers.

## 6.2. Matériels et méthodes

### 6.2.1. Jeu de données de calibration

Pour étudier le comportement écologique d'*Acer campestre* dans les forêts de France, nous avons extrait 3 286 relevés phytoécologiques d'EcoPlant, une base de données qui stocke des descriptions floristiques (relevés floristiques complets) avec des descriptions de milieu (variables liées au climat et au sol mesurées sur le terrain ou analysées chimiquement ou physiquement en laboratoire). Actuellement, EcoPlant stocke plus de 6 400 relevés forestiers répartis sur toute la France (Gégout et al. 2005), mais nous avons sélectionné 3 286 relevés qui présentaient suffisamment d'information édaphique, parmi lesquels *Acer campestre* était présent 460 fois (figure 6.35). La plupart des relevés de calibration présentent une superficie de 400 m<sup>2</sup>, en accord avec les pratiques courantes de la phytoécologie (Mueller-Dombois et Ellenberg 1974). Nous avons utilisé des données de présence/absence, où la valeur 1 représentait des relevés où *Acer campestre* était présent et la valeur 0 représentait des relevés où il était absent.



**Figure 6.35.** Localisation des 3 286 relevés forestiers étudiés en France (jeu de données de calibration), avec l'indication de la présence (460 points noirs) et de l'absence (2 826 points blancs) d'*Acer campestre* (source : EcoPlant, Gégout et al. 2005).

### 6.2.1.1. Données climatiques

Tous les 3 286 relevés sélectionnés étaient localisés avec une précision inférieure au kilomètre, et nous avons pu avoir accès à 159 variables et indices climatiques, disponibles sous forme de couches de résolution 1-km<sup>2</sup> dans un système d'information géographique (SIG) couvrant la France entière. Les variables climatiques ont été obtenues à l'aide d'un modèle numérique de terrain (MNT) et du modèle météorologique Aurelhy (Benichou et Le Breton 1987, tableau 6.6). Nous avons pris garde à sélectionner des variables climatiques qui avaient une influence sur (i) la croissance des plantes (T : températures moyennes, T05.09 : température moyenne entre mai et septembre, GDD6 : nombre de degrés-jour au-delà de 6 °C, VP6 : durée de la période de végétation au-delà de 6 °C), (ii) la production primaire (PET : évapotranspiration potentielle, AET : évapotranspiration réelle, R : radiations moyennes), (iii) les conditions de gel (Tn : températures minimales, FD : nombre de jours de gel) et (iv) les conditions de sécheresse (Tx : températures maximales, P : précipitations moyennes, MoistRat : indice d'humidité, SMD : déficit hydrique, SMS : surplus hydrique, DP : durée de la période de sécheresse, WB : bilan hydrique, voir le tableau 6.6 pour les définitions). Nous avons suivi deux voies différentes pour calculer l'évapotranspiration potentielle, en se basant sur les travaux de Thornthwaite et Mather (1957), et Turc (1961). La plupart de ces variables climatiques ont été testées avec succès dans des études récentes de modélisation de la distribution de plantes (par exemple Leathwick et Rogers 1996, Leathwick et Whitehead 2001, Pearson et al. 2002, Thuiller 2003). Bien que les variables climatiques révélaient une forte corrélation entre elles (et plus particulièrement les variables de mois successifs), notre but était de caractériser un jeu réduit de variables climatiques pertinentes pour décrire au mieux la distribution géographique d'*Acer campestre*.

### 6.2.1.2. Données édaphiques

De nombreux auteurs ont promu l'utilisation de variables édaphiques directes dans les études concernant l'écologie des végétaux (par exemple Austin et Meyers 1996, Austin 2002, Lawesson et Oksanen 2002, Diekmann 2003). En pratique cependant, ces variables ont rarement été intégrées dans les modèles, puisque l'accès à des variables de sol analytiques demande un prélèvement d'échantillon de sol sur le terrain et des analyses complémentaires en laboratoire. A la place, les caractéristiques édaphiques des placettes sont calculées à l'aide des relevés floristiques et des valeurs indicatrices d'Ellenberg associées (Ellenberg et al. 1992), qui représentent des valeurs numériques subjectives de la réponse des espèces aux conditions de milieu. En effet, Ellenberg et al. (1992) ont établi une liste de valeurs préférentielles pour l'humidité, le niveau d'acidité du sol, la disponibilité en éléments minéraux, la salinité, le régime de lumière, la température et la continentalité, ceci pour 2 726 plantes vasculaires de l'Europe centrale. Les valeurs indicatrices d'Ellenberg permettent l'estimation de variables environnementales locales bio-indiquées, lorsqu'un relevé floristique complet pour un site est disponible (ter Braak et Barendregt 1986). Dans ce chapitre cependant, pour les 3 286 relevés extraits d'EcoPlant, nous avons utilisé l'information disponible pour six variables édaphiques mesurées, représentant l'acidité (pH), les conditions d'azote (C/N), le taux de saturation (S/T) et les concentrations de trois éléments nutritifs (Ca, Mg et K sous leur forme logarithmique, tableau 6.6).

## 6.2.2. Jeu de données de validation

Pour valider nos modèles climatique et édaphique/climatique, nous avons eu accès à 88 004 relevés phytosociologiques indépendants provenant de l'Inventaire Forestier National (IFN), répartis sur toute la France, et pour lesquels nous avons l'information de présence/absence d'*Acer campestre* (figure 6.41a). Les régions pour lesquelles aucune information n'était disponible

sont représentées par des zones blanches dans la figure 6.41. La stratégie d'échantillonnage de l'Inventaire Forestier National (IFN) repose sur une grille de 500 m de côté sur toute la France, au sein de laquelle les points échantillonnés sont choisis de manière aléatoire. La distance minimale entre deux points d'évaluation est donc de 500 mètres.

Les valeurs des variables climatiques ont pu être obtenues par des manipulations sous système d'information géographique (SIG) pour ces 88 004 relevés, de la même manière que pour le jeu de données de calibration (relevés d'EcoPlant). Nous avons pu également calculer une valeur bio-indiquée pour trois des variables édaphiques sélectionnées (pH, S/T et C/N), comme la moyenne pondérée des optima écologiques de toutes les espèces présentes sur chaque relevé (ter Braak et Barendregt 1986). Les optima des espèces ont été extraits d'un catalogue non publié de valeurs indicatrices édaphiques pour 700 plantes forestières en France (Gégout et al. 2002, 2003). La plupart du temps du fait d'un manque de données, les variables édaphiques sont difficiles à cartographier à large échelle, et des cartes précises de caractéristiques édaphiques analytiques restent à développer. Pour cette étude, nous avons créé des cartes édaphiques de résolution 1 km<sup>2</sup> pour trois variables édaphiques (pH, C/N et S/T) par interpolation spatiale des valeurs bio-indiquées des 88 004 points du jeu de données de validation (« inverse distance weighting », voir Philip et Watson 1982).

### 6.2.3. Régression logistique

Nous avons utilisé la régression logistique pas-à-pas ascendante pour modéliser la réponse d'*Acer campestre* aux facteurs écologiques (ter Braak et Looman 1986, McCullagh et Nelder 1997). La régression logistique, cas particulier de modèle linéaire généralisé avec un lien logit et une distribution binomiale, a été souvent mise en œuvre pour caractériser les relations espèces/environnement (par exemple Odland et al. 1995), et est facile à implémenter dans la plupart des logiciels statistiques. Nous avons construit deux modèles différents pour *Acer campestre*, le premier avec seulement des variables climatiques (159 variables candidates), et le second avec des variables climatiques et édaphiques (165 variables candidates). Dans chaque cas, une procédure de sélection pas-à-pas ascendante, basée sur la déviance expliquée maximale, a été adoptée pour sélectionner les variables les plus pertinentes pour expliquer le comportement écologique de l'essence (McCullagh et Nelder 1997, Bio 2000). Les variables simples et leur carré ont été testées à chaque pas, afin de tenir compte de courbes de réponse potentiellement en forme de cloche (ter Braak et Looman 1986). Lorsque l'intégration d'une variable sous sa forme quadratique (testée par un test sur la déviance résiduelle selon McCullagh et Nelder 1997) n'était pas significative (au seuil de 0.001), la variable n'était gardée que sous sa forme simple, afin de pouvoir décrire une courbe de réponse monotone. La procédure pas-à-pas a été arrêtée lorsque que l'intégration d'une variable supplémentaire (sous sa forme simple ou combinée avec sa forme quadratique) ne conduisait pas à une réduction significative de la déviance résiduelle au seuil de 0.001.

**Tableau 6.6.** Liste des variables (a) climatiques et (b) édaphiques disponibles pour modéliser la distribution d'*Acer campestre* dans les forêts de France, avec indication des valeurs minimale, moyenne et maximale pour chaque variable.

<b>(a) Variables climatiques</b>					
Code	Variable	Unité	Minimum	Moyenne	Maximum
T	Température annuelle <sup>1</sup>	°C	4,8	9,7	13,5
T05.09	Température de mai à septembre	°C	10,8	15,7	19,3
GDD6	Degrés-jours (au-dessus de 6 °C)	°C	754	1748	2743
VP6	Durée de la période de végétation (au-dessus de 6 °C)	jours	163	246	365
Tn	Température minimale annuelle <sup>1</sup>	°C	-0,2	5,0	9,2
FD	Jours de gel <sup>1</sup>	jours	21	77	180
Tx	Température maximale annuelle <sup>1</sup>	°C	9,5	14,4	18,8
DP	Durée de la période sèche	jours	0	129	223
R	Radiations annuelles <sup>1</sup>	kJ/cm <sup>2</sup>	294	429	590
P	Précipitations annuelles <sup>1</sup>	mm	542	990	2264
pAu	Précipitations automnales (oct., nov., déc.) <sup>3</sup>	mm	111	280	714
PET	Evapotranspiration potentielle <sup>1,2</sup>	mm	435	673	912
AET	Evapotranspiration réelle <sup>1,2</sup>	mm	378	509	685
AETTh	Evapotranspiration réelle selon Thornthwaite <sup>3</sup>	mm	402	502	611
MoistRat	Indice d'humidité (1-AET/PET) <sup>1,2</sup>	-	0,00	0,23	0,47
SMD	Déficit hydrique annuel (AET-PET) <sup>1,2</sup>	mm	-399	-163	0
SMS	Surplus hydrique annuel (P-PET) <sup>1,2</sup>	mm	90	476	1685
WB	Bilan hydrique (somme des P-PET mensuels) <sup>1,2</sup>	mm	-218	312	1682

<sup>1</sup> Les valeurs mensuelles étaient également disponibles et ont été testées dans les modèles. <sup>2</sup> Deux méthodes différentes ont été utilisées pour calculer l'évapotranspiration potentielle, en se basant sur les travaux de (i) [Turc \(1961\)](#), et (ii) [Thornthwaite et Mather \(1957\)](#). Par défaut, les chiffres présentés dans ce tableau ont été calculés avec la méthode de [Turc \(1961\)](#). <sup>3</sup> pAu et AETTh ont été ajoutés à ce tableau puisque ces variables contribuent à expliquer la distribution d'*Acer campestre* (voir tableau 6.7).

<b>(b) Variables édaphiques</b>					
Code	Variable	Unité	Minimum	Moyenne	Maximum
pH	pH eau du sol	-	3,3	5,1	8,2
ST	Taux de saturation en bases échangeables	%	0,0	51,6	100,0
CN	Rapport C/N	-	7,9	17,0	45,8
lnCa	Concentration en calcium <sup>4</sup>	mEq	-3,2	0,8	4,4
lnMg	Concentration en magnésium <sup>4</sup>	mEq	-4,2	-0,7	2,3
lnK	Concentration en potassium <sup>4</sup>	mEq	-3,9	-1,4	0,4

<sup>4</sup> Les concentrations en calcium, magnésium, et potassium sont exprimées en mEq pour 100 g de sol sec, mais sont présentées sous leur forme logarithmique.

#### 6.2.4. Evaluation quantitative des modèles

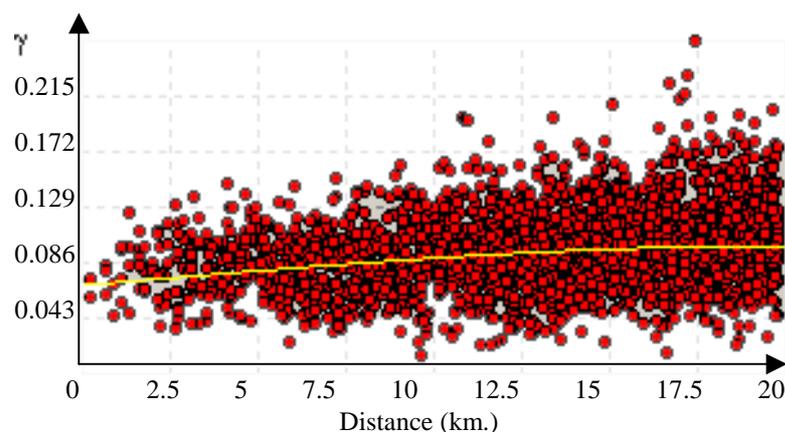
Pour comparer les probabilités de présence d'*Acer campestre* résultant des deux modèles, avec les vraies données de présence/absence, nous avons sélectionné un seuil optimal de probabilité à l'aide d'une courbe ROC (receiver-operating characteristic, [Fielding et Bell 1997](#)). Ce seuil a permis la transformation des valeurs de probabilités de présence en données binaires de présence/absence. Nous avons calculé une matrice de confusion pour ce seuil. De la matrice de confusion, nous avons calculé trois mesures synthétiques pour qualifier le taux de classifications correctes des présences/absences observées et prédites : (i) taux de succès global (S, taux de bonne prédiction de la présence et de l'absence), (ii) sensibilité (Sn, taux de bonne prédiction de la

présence), (iii) spécificité (Sp, taux de bonne prédiction de l'absence). Nous avons utilisé l'aire sous la courbe ROC pour évaluer la qualité de prédiction des modèles (Fielding et Bell 1997, Elith 2002, Elith et Burgman 2002). Tous les calculs ont été effectués avec le logiciel S-Plus (MathSoft 1999).

### 6.2.5. Autocorrélation spatiale

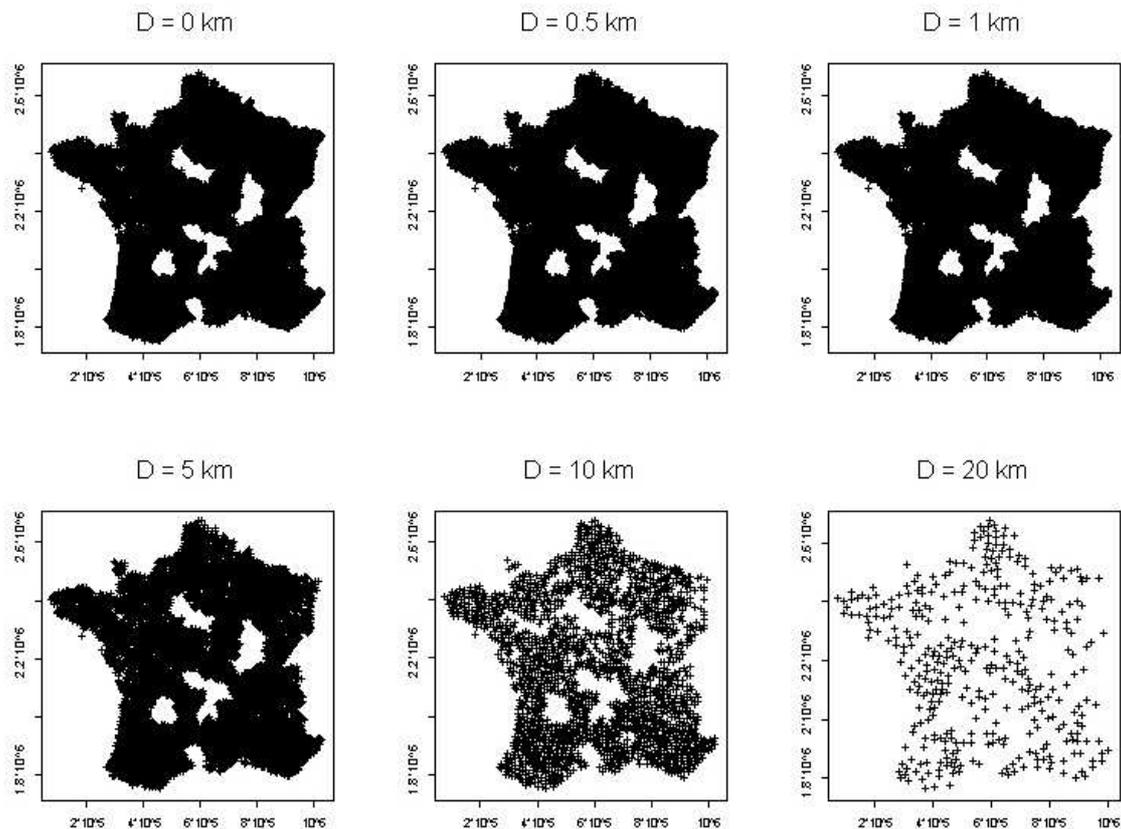
L'autocorrélation spatiale peut augmenter artificiellement le pouvoir prédictif des modèles (Legendre 1993, Lennon 2000, Hampe 2004, Araújo et al. 2005), si la distance entre les relevés (de calibration ou d'évaluation) est inférieure à la distance minimale pour laquelle on n'observe plus d'autocorrélation. Pour éviter ce problème, nous avons sélectionné des relevés de calibration qui étaient géographiquement ou écologiquement indépendants les uns des autres (voir le premier chapitre). Nous avons utilisé la règle suivante : lorsque deux relevés étaient séparés de moins de 500 mètres, nous ne gardions les deux relevés que si leurs conditions écologiques différaient en terme d'altitude (au moins 50 m), d'exposition (au moins 90 °) ou de pH du sol (au moins une unité).

Nous avons utilisé une technique géostatistique particulière (« indicator kriging », Journel 1983, Isaaks et Srivastava 1989), qui est adaptée aux données catégorielles telles que la présence/absence, et qui est basée sur le krigeage ordinaire (voir Bayliss et al. 2005 pour une introduction de la technique en écologie). Dans ce chapitre, la technique « indicator kriging » a été utilisée directement sur les données de présence/absence du jeu d'évaluation (voir également Miller et Franklin 2002, Marinoni 2003, Bayliss et al. 2005). Un semi-variogramme a ensuite été calculé à l'aide d'ArcGis (ESRI 2002) pour évaluer la dépendance spatiale ( $\gamma$ ) en fonction de la distance entre les relevés d'évaluation. Nous avons examiné cette relation pour de courtes distances (inférieures à 20 km), car d'autres études ont montré que l'abondance des Erables (*Acer* sp.) pouvait révéler une certaine autocorrélation à l'échelle locale (par exemple Clark et al. 1998, Schwarz et al. 2003). En effet, la présence d'un arbre peut augmenter la probabilité de présence d'arbres voisins dans le rayon de dispersion de l'espèce. Cependant, aucune augmentation sensible n'a pu être observée sur le semivariogramme pour de faibles distances (voir la figure 6.36 pour les 88 004 points d'évaluation), ce qui signifie que la dispersion de l'espèce ne semble pas avoir d'effet important sur l'autocorrélation spatiale à faible distance.



**Figure 6.36.** Semivariogramme calculé avec la technique « indicator kriging » sur les données de présence/absence d'*Acer campestre*, pour le jeu de données d'évaluation (88 004 relevés). Chaque point représente des paires de points qui ont une distance commune (taille du « lag ») et une direction commune entre eux (ESRI 2002). 40 « lags » de 500 mètres sont représentés, révélant ainsi les relations entre les relevés jusqu'à une distance de 20 km.

Afin d'évaluer plus profondément l'influence de l'autocorrélation spatiale sur le pouvoir prédictif de nos modèles, les quatre mesures d'évaluation (AUC : aire sous la courbe ROC, S : le succès global, Sn : la sensibilité et Sp : la spécificité) ont été calculées pour différents extraits du jeu de données d'évaluation. Tout d'abord, des jeux d'évaluation ont été sélectionnés en les séparant des relevés de calibration de distances croissantes : 500 m, 1 km, 5 km, 10 km et 20 km. Ensuite, des jeux d'évaluation ont été sélectionnés en les séparant des relevés de calibration et d'eux-mêmes de ces mêmes distances croissantes (figure 6.37). Finalement, des jeux d'évaluation de taille variable ont été sélectionnés au hasard.



**Figure 6.37.** Cartes des jeux d'évaluation successifs avec une distance (D) croissante entre les points d'évaluation et entre les points d'évaluation et les points de calibration. Les nombres de points pour les distances égales à 0, 0,5, 1, 5, 10 et 20 km sont respectivement 88 004, 81 554, 54 879, 8 359, 2 2278 et 374.

## 6.3. Résultats

### 6.3.1. Réponse climatique d'*Acer campestre*

La procédure de sélection pas-à-pas ascendante a conduit à un modèle climatique à deux variables, incluant les précipitations automnales moyennes (pAu, somme des précipitations moyennes d'octobre, novembre et décembre), et l'évapotranspiration réelle moyenne calculée avec la méthode de Thornthwaite. L'équation finale liant la probabilité de présence d'*Acer campestre* aux facteurs climatiques était :

$$\ln\left(\frac{p}{1-p}\right) = -6.331 - 0.01462 \cdot pAu + 0.01636 \cdot AETTh \quad \text{(Equation 6.9)}$$

La réponse d'*Acer campestre* aux deux variables était monotone, soit décroissante pour les précipitations automnales (préférence de l'espèce pour des stations sèches), ou croissante pour l'évapotranspiration réelle selon Thornthwaite (préférence de l'espèce pour des conditions de croissance maximales). La réduction de déviance associée aux deux variables climatiques était comparable (tableau 6.7), et la surface de réponse climatique du modèle climatique à deux variables est présentée figure 6.38. L'inclusion d'une variable climatique supplémentaire à ce modèle n'a pas apporté d'amélioration significative en termes de réduction de déviance.

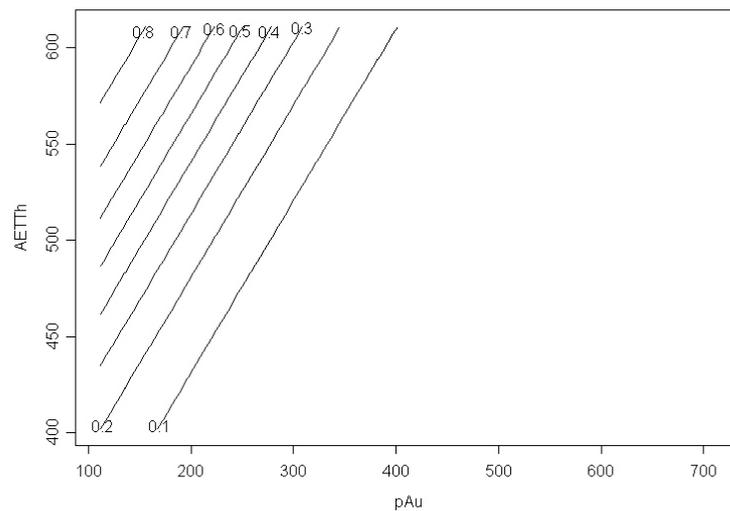
**Tableau 6.7.** Résumé du modèle (a) climatique et (b) climatique et édaphique, calculés par régression logistique pas-à-pas ascendante pour lier la présence/absence d'*Acer campestre* dans les forêts de France avec les variables écologiques disponibles. Des informations sont données à chaque étape du calcul sur la variable incluse au modèle sous sa forme quadratique ou linéaire. D2 est calculé par (déviance nulle - déviance du modèle)/déviance nulle. La valeur AUC est l'aire sous la courbe ROC, et S, Sn et Sp (%) sont calculés à partir de la matrice de confusion et représentent respectivement le taux de succès global, le taux de bonnes prédictions de la présence et le taux de bonnes prédictions de l'absence (Fielding et Bell 1997).

<b>(a)</b>		Déviance nulle = 2 661				Calibration (n = 3 286)			Validation (n = 88 004)				
Etape	Variable ajoutée	Quadratique (O/N)	Déviance du modèle	Réduction de déviance	D2	AUC	S	Sn	Sp	AUC	S	Sn	Sp
1	pAu	N	2543	118	0,044	0,64	54	69	52	0,56	38	81	30
2	AETTh	N	2427	235	0,088	<b>0,72</b>	<b>55</b>	<b>84</b>	<b>50</b>	<b>0,64</b>	<b>54</b>	<b>73</b>	<b>50</b>

<b>(b)</b>		Déviance nulle = 2 661				Calibration (n = 3 286)			Validation (n = 88 004)				
Step	Variable ajoutée	Quadratique (O/N)	Déviance du modèle	Réduction de déviance	D2	AUC	S	Sn	Sp	AUC	S	Sn	Sp
1	pH	Y	1933	728	0.273	0.85	76	84	75	0.83	71	88	67
2	pAu	N	1835	826	0.310	0.88	73	92	70	0.81	66	89	62
3	AETTh	N	1744	917	0.344	<b>0.89</b>	<b>81</b>	<b>85</b>	<b>81</b>	<b>0.85</b>	<b>73</b>	<b>86</b>	<b>70</b>

La transposition du modèle climatique (équation 6.9) dans l'espace géographique est présentée figure 6.41c. Le modèle prédit de faibles probabilités de présence d'*Acer campestre* dans les régions montagnardes (figure 6.41b, régions 1a : Vosges, 1b : Jura, 1c : Alpes, 1d : Massif central, et 1e : Pyrénées) et hyper-atlantiques (par exemple figure 6.41b, région 3b : Bretagne), qui sont des zones où les précipitations automnales semblent être trop élevées pour favoriser la présence d'*Acer campestre*. Le modèle prédit également de faibles probabilités de présence dans la zone méditerranéenne (figure 6.41b, région 2), où les précipitations automnales sont favorables, mais où l'évapotranspiration réelle selon Thornthwaite est faible. Le modèle climatique prédit de plus grandes probabilités de présence d'*Acer campestre* dans une zone sud-ouest/nord-est excluant les zones de montagnes. Néanmoins, la qualité médiocre (*sensu* Swets 1988) de ce modèle est illustrée par une faible valeur de l'aire sous la courbe ROC (AUC = 0,72 pour le jeu de données de calibration, et AUC = 0,64 pour le jeu de données de validation, voir la figure 6.38 et le tableau 6.7a).



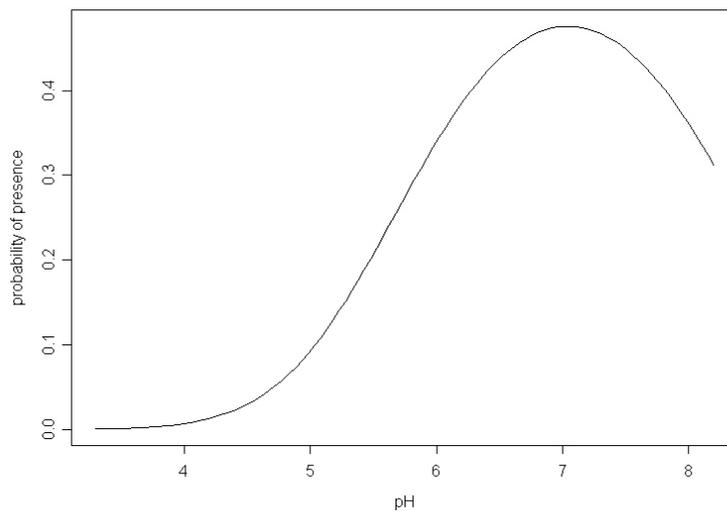
**Figure 6.38.** Surface de réponse climatique d'*Acer campestre*, vis-à-vis des précipitations moyennes automnales (pAu) et de l'évapotranspiration réelle selon Thorthwaite (AETTh), calculée par la régression logistique sur 3 286 relevés forestiers. Voir le tableau 6.7 pour plus de détails sur le modèle climatique.

### 6.3.2. Réponse climatique et édaphique d'*Acer campestre*

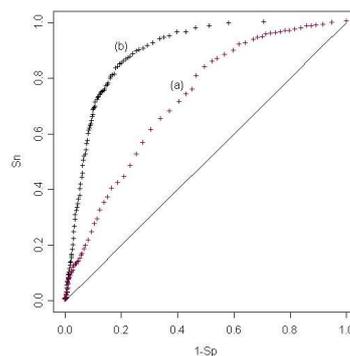
Chacune des six variables édaphiques était significativement meilleure que toutes les variables climatiques en terme de réduction de déviance, dès la première étape de la procédure de sélection pas-à-pas. Deux variables édaphiques parmi les six étaient les meilleures : lnCa et pH, mais nous avons choisi de conserver la variable pH dans le modèle final, parce que cette variable ne présentait aucune valeur manquante (contrairement à lnCa, qui présentait 314 valeurs manquantes). Le pH est un gradient environnemental complexe, qui est facile à relever sur le terrain, et qui est souvent utilisé dans les études de relations espèces/environnement (voir par exemple Coudun et Gégout 2005). Deux pas supplémentaires ont conduit à inclure des variables climatiques dans le modèle, avec les deux variables du modèle climatique précédent parmi les meilleures dans notre procédure de sélection (tableau 6.7b). Nous avons donc choisi de conserver les précipitations automnales (pAu) et l'évapotranspiration réelle selon Thorthwaite (AETTh) dans notre modèle climatique et édaphique final, donc l'équation logistique est :

$$\ln\left(\frac{p}{1-p}\right) = -30.47 + 7.56 \cdot \text{pH} - 0.536 \cdot \text{pH}^2 - 0.01235 \cdot \text{pAu} + 0.01378 \cdot \text{AETTh} \quad \text{(Equation 6.10)}$$

La nature calcicole d'*Acer campestre* était évidente (valeurs de pH optimales comprises entre 6,5 et 7,5 pour l'espèce), en considérant sa courbe de réponse vis-à-vis du pH (figure 6.39). Le comportement climatique d'*Acer campestre* restait stable lorsque nous avons incorporé le pH parmi les variables explicatrices.



**Figure 6.39.** Courbe de réponse écologique d'*Acer campestre* vis-à-vis du pH du sol, calculé par régression logistique sur 3 286 relevés forestiers. Voir le tableau 6.7b pour plus de détails sur le modèle climatique et édaphique.



**Figure 6.40.** Courbes ROC associées (a) au modèle climatique, (b) au modèle climatique et édaphique.

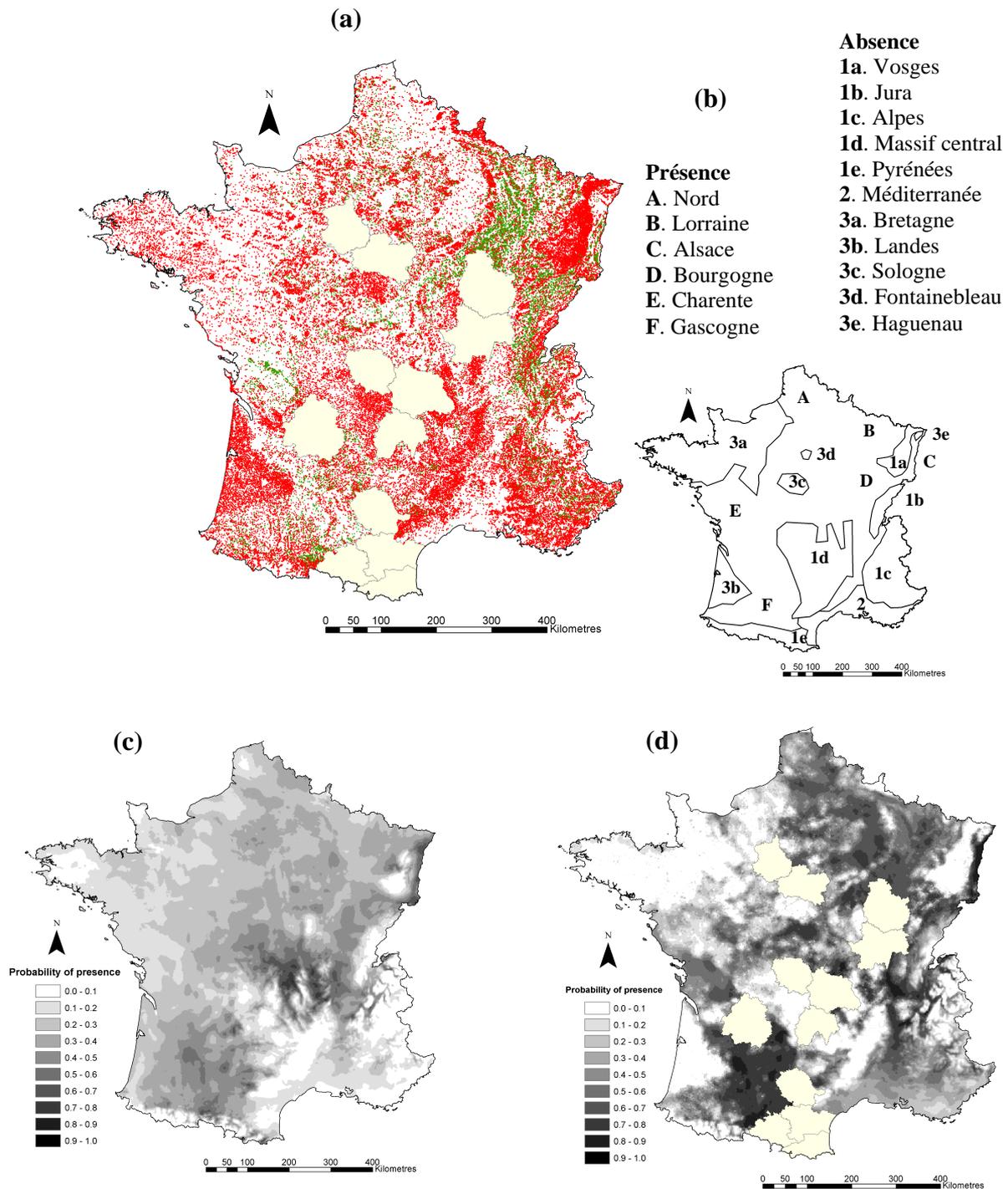
La transposition du modèle climatique et édaphique (équation 6.10) dans l'espace géographique est représentée par la figure 6.41d. Ce modèle exclut encore *Acer campestre* des zones montagnardes et méditerranéennes (figure 6.41b, régions 1a : Vosges, 1b : Jura, 1c : Alpes, 1d : Massif central, 1e : Pyrénées, et 2 : Méditerranée) avec de faibles probabilités de présence prédites. Ce modèle exclut également l'espèce des régions acides, où elle ne peut pas croître et survivre (figure 6.37). Ainsi, la prédiction de l'absence de l'espèce est bien meilleure qu'avec le modèle climatique, et permet donc une valeur plus forte de la spécificité (tableau 6.7). Par exemple, l'absence d'*Acer campestre* est prédite des zones acides cristallines primaires (figure 6.41b, région 3a : Bretagne), et des zones acides sableuses du Tertiaire/Quaternaire (figure 6.41b, régions 3b : Landes et 3c : Sologne). A des échelles locales, de très faibles probabilités de présence sont prédites par exemple pour les forêts acides de Fontainebleau (figure 6.41b, région 3d) ou Haguenau (figure 6.41b, région 3e). Avec des probabilités de présence prédites fortes dans le Nord, en Lorraine, Alsace, Bourgogne, Charente, et Gascogne (figure 6.41b, respectivement régions A, B, C, D, E, et F), le modèle est également capable de prédire correctement la présence

d'*Acer campestre* en France. La bonne qualité du modèle (*sensu* Swets 1988) est illustrée par une forte valeur de l'aire sous la courbe ROC (AUC = 0,89 pour le jeu de données de calibration et AUC = 0,85 pour le jeu de données de validation, figure 6.38 et tableau 6.7b).

Le pouvoir prédictif du modèle climatique ou du modèle climatique et édaphique ne semblait pas varier beaucoup lorsque les points d'évaluation étaient sélectionnés de plus en plus loin des points de calibration (tableau 6.8a). Plus de 98 % des points d'évaluation sont au moins séparés de 500 m d'un point de calibration. Avec un taux de succès global  $S = 71\%$ , le modèle restait valide lorsque tous les points d'évaluation étaient séparés d'au moins 20 km des points de calibration d'EcoPlant (tableau 6.8a). Cela correspond géographiquement par exemple à la région Charente (region E, figure 6.41b) où aucun relevé n'était disponible pour calibrer le modèle (figure 6.35), mais où la carte de prédiction s'ajuste relativement bien aux données d'observation. Cependant, lorsque les points d'évaluation étaient sélectionnés de plus en plus loin des points de calibration et d'eux-mêmes, une baisse du pouvoir prédictif pouvait être remarquée (Tableau 6.8b). Enfin, lorsque les points d'évaluation ont été sélectionnés au hasard, le pouvoir prédictif du modèle restait comparable au pouvoir prédictif calculé avec la totalité du jeu de données d'évaluation (tableau 6.8c et 6.8c).

**Tableau 6.8.** Pouvoir prédictif des modèles de distribution d'*Acer campestre* calculé avec des variables climatiques et édaphiques, en utilisant des jeux d'évaluation obtenus : (a) en augmentant la distance minimale entre les points d'évaluation et les points de calibration ; (b) en augmentant la distance minimale entre les points d'évaluation et entre les points d'évaluation et les points de calibration ; (c) en sélectionnant au hasard des points d'évaluation. Le pouvoir prédictif calculé avec le jeu d'évaluation complet est rappelé en (d). Nplots est le nombre de points d'évaluation sélectionnés, AUC est l'aire sous la courbe ROC et S, Sn et Sp sont respectivement le succès global, la sensibilité et la spécificité (en %).

	Nplots	Npres	AUC	S	Sn	Sp
<b>(a) distance minimale entre les points d'évaluation et les points de calibration</b>						
0.5 km	86 825	14 251	0,84	73	86	71
1 km	84 373	13 956	0,84	73	86	71
5 km	65 278	11 270	0,83	71	86	68
10 km	47 756	7 610	0,82	70	85	67
20 km	13 025	2 220	0,82	71	88	67
<b>(b) distance minimale entre les points d'évaluation et entre les points d'évaluation et les points de calibration</b>						
0.5 km	81 554	13 630	0,84	73	86	71
1 km	54 879	9 696	0,83	71	88	67
5 km	8 359	1 463	0,80	67	88	62
10 km	2 278	393	0,79	69	83	67
20 km	374	65	0,82	66	92	61
<b>(c) selection au hasard des points d'évaluation</b>						
-	81 554	13 339	0,85	73	86	71
-	54 879	8 919	0,85	73	86	70
-	8 359	1 459	0,84	75	84	73
-	2 278	392	0,85	70	93	65
-	374	73	0,84	69	95	62
<b>(d) jeu de données d'évaluation complet</b>						
-	88 004	14 373	0,85	73	86	70



**Figure 6.41.** Cartes de la présence/absence réelle d'*Acer campestre* et des probabilités de présence : (a) localisation des 88 004 relevés de l'Inventaire Forestier National (jeu de validation), avec indication de la présence (points verts) et de l'absence (points rouges) de l'espèce ; (b) schéma de la France avec indication de régions naturelles ou de forêts ; (c) carte des probabilités de présence prédites par le modèle climatique (équation 6.9, figure 6.26, et tableau 6.7a), (d) carte des probabilités de présence prédites par le modèle climatique et édaphique (équation 6.10, figure 6.37, et tableau 6.7b). Les zones blanches sur les figures 6.41a et 6.41d représentent des départements français pour lesquels l'information floristique n'est pas encore disponible, et donc pour lesquels nous n'avons pas pu calculer de variable édaphique sur la base de la composition floristique des relevés.

## 6.4. Discussion

### 6.4.1. Remarques méthodologiques

La régression logistique s'est montrée être une technique robuste pour prédire la distribution d'*Acer campestre* dans les forêts de France, en accord avec d'autres études (Brotons et al. 2004, Segurado et Araújo 2004). Le nombre d'occurrences de l'espèce dans nos jeux de données était suffisant pour s'assurer d'une caractérisation fiable des relations entre l'espèce et son environnement (Coudun et Gégout en révision mineure). Même si cette technique contraint la forme des courbes de réponse le long de chaque gradient à être soit, plate, sigmoïde ou unimodale en forme de cloche, nous avons trouvé des bonnes valeurs d'AUC pour les jeux de données de calibration et de validation (*sensu* Swets 1988), avec un nombre parcimonieux de variables explicatrices. La caractérisation du comportement écologique d'*Acer campestre* s'est montrée également robuste, puisque sa réponse climatique semblait être stable lorsqu'une variable édaphique était introduite dans le modèle (équations 6.9 et 6.10, tableau 6.7). La stabilité du pouvoir prédictif du modèle lorsque la distance entre les points d'évaluation et les points de calibration était augmentée, supporte son caractère robuste.

Les techniques de modélisation de la distribution des espèces ou des communautés sont en cours d'amélioration, et plus particulièrement dans le contexte forestier (voir Maggini et al. sous presse). Pour améliorer les modèles écologiques, plus de relevés seraient nécessaires dans la base de données EcoPlant, et plus particulièrement dans la région méditerranéenne. Pour améliorer les cartes de distribution, la base de données de l'Inventaire Forestier National (IFN) a besoin d'être complétée (figure 6.41a). En tous les cas, cette étude s'est appuyée sur deux grosses bases de données nationales de relevés forestiers (Gégout et al. 2005), qui fournissent ensemble des données de plus grande qualité que celles disponibles habituellement pour modéliser la distribution des espèces végétales. Par exemple, ces données sont plus adéquates que des données fournies par des atlas de distribution puisque des données écologiques locales et de l'information édaphique précise sont associées à chaque relevé floristique (Austin soumis).

### 6.4.2. Autocorrélation spatiale

L'autocorrélation spatiale peut biaiser les modèles de distribution basés sur le concept de la niche écologique (Diniz-Filho et al. 2003, Munoz et Felicísimo 2004, Segurado et al. soumis). Dans cette étude, l'autocorrélation spatiale due à la dispersion de l'espèce n'influence pas le succès des prédictions, probablement parce que la distance entre les relevés (supérieure à 1 km pour plus de 95 % des relevés) est plus grande que la distance de dispersion de l'espèce. Des études sur les patrons locaux de distribution des espèces d'Erables ont montré que la distance maximale de dispersion est souvent inférieure à 0,5 km (Guries et Nordheim 1984, Clark et al. 1998, Schwarz et al. 2003). La distance de dispersion d'*Acer campestre* n'a pas été étudiée dans ce chapitre, mais la faible hauteur de cette espèce (jusqu'à 15 m, Rameau et al. 1989) limite probablement sa dispersion, comme c'est le cas pour d'autres espèces d'Erable (Guries et Nordheim 1984).

La fragmentation de la présence/absence à l'échelle locale est probablement due à l'autocorrélation spatiale des variables écologiques prédictives : il est relativement probable de trouver des conditions écologiques similaires dans des relevés proches. L'augmentation de la distance entre les points d'évaluation et entre les points d'évaluation et de calibration a conduit à l'élimination progressive des points d'évaluation rencontrés à l'intérieur de larges régions forestières en France (telles que le nord-est et le sud-ouest de la France ou les zones de montagne), et à la conservation de points d'évaluation dans des petites régions forestières (surtout dans des régions agricoles telles que le nord-ouest de la France). Avec de grandes

distances minimales entre les relevés (10 et 20 km, tableau 6.8), les points d'évaluation ont été sélectionnés surtout dans des petites zones forestières isolées, sur-représentant les petites forêts vis-à-vis de la totalité de la superficie forestière de France. Ceci explique probablement la baisse de pouvoir prédictif observée (tableau 6.8b). Ceci est également confirmé par un pouvoir prédictif similaire lorsque tout le jeu d'évaluation est sélectionné ou lorsque les relevés sont tirés au hasard parmi le jeu complet (tableau 6.8c). En effet la sélection au hasard assure une probabilité d'échantillonner un relevé d'évaluation proportionnelle à la superficie de la zone forestière sur laquelle le relevé est présent.

#### 6.4.3. L'importance des variables nutritionnelles du sol

La plupart des études récemment publiées sur la modélisation de la distribution d'espèces végétales n'utilisaient que des variables climatiques (par exemple Vetaas 2002), mais notre étude a montré qu'*Acer campestre* était une espèce particulièrement pertinente pour illustrer le besoin d'intégrer des facteurs édaphiques directs dans les modèles de prédiction de la distribution de certaines plantes terrestres. Même si certaines études intègrent des variables liées au sol ou au relief (topographie, lithologie, etc.), elles n'utilisent pas de variables directes mesurées (Le Duc et al. 1992, Firbank et al. 1998, Franklin 1998, 2002, Franklin et al. 2000, Fertig et Reiners 2002), qui ont un impact physiologique sur la croissance des plantes. Contrairement aux travaux de Huntley et al. (2004), nous montrons que l'influence des facteurs climatiques sur *Acer campestre* est bien plus faible que l'influence des conditions de pH du sol. Des études complémentaires devraient permettre de quantifier l'importance relative des facteurs climatiques et des facteurs édaphiques pour d'autres plantes forestières, en utilisant par exemple la procédure décrite par Zimmermann et al. (soumis).

Le choix de se focaliser sur une seule espèce (*Acer campestre*) en France était gouverné par la volonté d'illustrer que les facteurs édaphiques, lorsque des données de bonne qualité sont disponibles, peuvent jouer un rôle important dans la modélisation de la distribution des espèces végétales. Nous avons étudié le comportement d'*Acer campestre* dans une partie représentative de son aire de répartition en Europe, puisque la gamme des conditions écologiques rencontrées en France est large (tableau 6.6). La préférence d'*Acer campestre* pour des fortes valeurs d'évapotranspiration réelle limite la gamme de conditions écologiques dans lesquelles l'espèce peut survivre puisque les conditions ne doivent être ni trop froides (climat de montagne), ni trop sèches (climat méditerranéen). La forte probabilité de présence de l'espèce dans des régions de faibles précipitations automnales est cohérente avec l'idée que l'espèce a une faible résistance envers des climats trop pluvieux, par exemple, l'espèce est absente de l'Ecosse et de l'Irlande, qui sont des régions particulièrement pluvieuses en Europe.

La prédiction de sa distribution avec des variables climatiques seulement n'a donné qu'une estimation grossière de son aire de répartition à l'échelle du pays, et l'incorporation du pH du sol (cohérente avec les observations des écologistes comme Rameau et al. 1989 ou Ellenberg et al. 1992), agissant comme un filtre local, a permis une prédiction plus fine de la présence et de l'absence de l'espèce. La validité des modèles prédictifs de distribution des plantes doit être encore être testée dans différentes régions (Schröder et Richter 1999, Gégout et Krizova 2003, Coudun et Gégout 2005), et à différentes échelles spatiales (Cherrill et al. 1995, Collingham et al. 2000, Lipsett-Moore et al. 2003, Thuiller et al. 2003a). Le modèle national de distribution d'*Acer campestre* a par exemple été comparé récemment à un modèle régional dans les Alpes françaises (Py 2005), soulignant encore l'intérêt de travailler à différentes échelles. En effet, des cartes de distribution prédite, valides aux échelles nationales et/ou locales, présentent un grand intérêt pour la gestion des espèces et leur conservation (Fortin et al. 2005), et pourraient être déterminées pour d'autres espèces forestières en France de la même manière que celle présentée dans cette étude.

## Conclusions et perspectives



L'analyse quantitative du comportement écologique des espèces à l'échelle de la France a été menée dans ce travail pour explorer l'apport des grosses bases de données possédant simultanément des données floristiques, climatiques et édaphiques, dans le but d'approfondir la connaissance de la niche écologique des espèces végétales forestières.

Ce travail a été conduit à partir d'une sélection de relevés de la base de données EcoPlant développée à l'ENGREF (Gégout et al. 2005). Les 6 432 placettes intégrées à ce jour à cette base couvrent la majeure partie de la gamme de valeurs observables en France, pour les principaux facteurs qui affectent la survie ou le développement des espèces. D'un point de vue climatique, pour des altitudes qui varient de 0 à 2 300 m, la gamme de températures moyennes annuelles s'étend de 3,5 à 15,5 °C et la gamme de précipitations moyennes annuelles s'étend de 450 à 2 400 mm. L'ensemble de la gamme des conditions observables dans la nature pour le pH, les ressources minérales en bases échangeables, ou l'azote, est également échantillonné.

Cette diversité de conditions de milieu rencontrées ouvre la possibilité, à la différence des études locales, d'étudier le comportement des plantes et de positionner précisément leur optimum et leur amplitude écologiques par rapport à une vaste gamme de valeurs des principaux facteurs environnementaux. Il est ainsi possible de connaître avec précision les valeurs des facteurs (climatiques ou édaphiques) pour lesquelles l'espèce apparaît dans les conditions naturelles, présente son optimum, et disparaît. Les données d'EcoPlant permettent ainsi de connaître le comportement des espèces sur l'ensemble de la gamme de ressources nutritionnelles présentes dans la nature et sur la gamme des conditions climatiques à caractère océanique des forêts tempérées de plaine et de montagnes européennes.

La base de données EcoPlant reste cependant incomplète à ce jour, puisque les typologies des stations forestières ne couvrent qu'une partie du territoire français. En effet, plusieurs régions sont actuellement encore peu étudiées, parmi lesquelles, (i) les montagnes françaises (aucune étude dans les Alpes internes, la partie sud des Pyrénées, une partie du Massif central, la haute chaîne du Jura), (ii) la région méditerranéenne, et (iii) certaines régions atlantiques (Aquitaine, Poitou-Charentes, Pays-de-la-Loire, Nord-Pas-de-Calais).

L'échantillonnage non aléatoire des relevés dans EcoPlant le long des gradients écologiques et dans l'espace géographique pourrait être à l'origine de biais liés à la sur-représentation de certaines localités comprenant de très nombreux relevés, comme c'est le cas par exemple pour le nord-est de la France. Ces phénomènes peuvent probablement modifier la courbe de réponse des espèces en perturbant l'effet général d'un facteur par des phénomènes d'autocorrélation spatiale (Legendre 1993, Koenig 1999). Dans nos traitements, ces biais ont été évités par un processus rigoureux de sélection des placettes utilisées dans les analyses. Ainsi, dans toute la thèse, nous avons procédé à une sélection écologique et géographique avant chaque étude, en ne conservant que des relevés séparés de plus de 500 mètres les uns des autres ou des relevés séparés de moins de 500 mètres mais présents dans des conditions stationnelles différentes. Cette sélection limite très certainement les biais qui pourraient être introduits dans la construction des courbes de réponse des espèces et le calcul de leur optimum et de leur amplitude écologiques.

## Intérêt d'un grand nombre de placettes pour caractériser le comportement écologique des espèces

L'analyse du nombre de relevés nécessaire pour déterminer de façon fiable le comportement écologique des espèces (évolution de la probabilité de présence des espèces avec les valeurs des gradients environnementaux, optimum, gamme de valeurs où l'espèce est potentiellement présente) n'avait été que très peu abordée par le passé. Les simulations, validées grâce à EcoPlant, montrent dans le chapitre 3 que les courbes de réponse élaborées avec moins de 100 placettes conduisent à des erreurs importantes dans l'évaluation de l'optimum et de l'amplitude écologiques des espèces, la plus importante étant une absence de réaction apparente alors que l'espèce réagit réellement au facteur. Pour des gammes de 100 à 1 000 relevés, seul le comportement écologique des espèces suffisamment fréquentes et à préférence écologique marquée peuvent être étudiées sans risque important d'erreurs. Plus précisément, notre travail montre que c'est essentiellement le nombre d'occurrences des espèces qui conditionne la possibilité ou non de déterminer leur réponse précise aux facteurs du milieu.

Les simulations montrent qu'un seuil de 50 à 100 occurrences d'une espèce semble être un minimum en dessous duquel il est difficile d'estimer avec précision les trois paramètres synthétiques de la courbe de réponse que sont l'optimum, l'amplitude et la probabilité maximale de présence (chapitre 3). Actuellement, parmi les 2 154 espèces rencontrées dans les 6 432 relevés intégrés à EcoPlant, 450 espèces seulement sont présentes plus de 50 fois, parmi lesquelles 279 espèces sont présentes plus de 100 fois. Si l'on souhaite connaître l'écologie de la majorité des espèces de la flore forestière française, le nombre de relevés d'EcoPlant doit être augmenté, et en particulier dans les zones mal connues présentées ci-dessus.

## Validité du comportement écologique à l'échelle de la France

Plusieurs études internationales et les concepts sous-jacents à l'élaboration régionale des typologies de stations forestières en France (Becker et Le Goff 1988, Brêthes 1989) laissaient penser que des différences importantes de comportement écologique des espèces pouvaient exister entre les régions françaises.

La comparaison de la réponse des espèces au pH de l'horizon A du sol en plaine entre le nord-est et le nord-ouest de la France, et en montagne entre les Vosges et le Jura, met principalement en évidence des changements de comportement limités d'une région française à une autre. Lorsqu'ils existent, ces changements sont liés à des différences de contexte de compétition entre régions étudiées : les espèces "acidiphiles" ont des optima plus élevés dans les régions où la compétition interspécifique est la plus faible dans les milieux favorables d'un point de vue nutritionnel.

Cette stabilité du comportement nutritionnel a pour conséquence importante d'ouvrir la voie à des études qui s'étendent sur de vastes territoires et dont la validité n'est pas remise en cause lorsque l'on change de région. Elle permet également l'analyse du comportement écologique des espèces sur des portions de gradients les plus larges possibles, plutôt que sur des portions limitées.

Parmi les trois principaux mécanismes pouvant conduire à un changement de comportement écologique dans des régions différentes (différentiation d'écotypes, compétition, compensation de facteurs), c'est la compétition qui a semblé être la cause des quelques changements observés dans le comportement des espèces. La préférence des espèces pour des milieux dont le pH se situe aux alentours de 6 à 7 a été montrée par trois fois (chapitres 3, 4, et 5) avec une forte densité

d'espèces dont l'optimum de pH se trouve dans cette gamme. Nous avons pu mettre en évidence des phénomènes de compétition différents dans ces milieux favorables entre le nord-est et le nord-ouest de la France (chapitre 4, Coudun et Gégout 2005), et dans une moindre mesure entre les Vosges et le Jura (chapitre 5), et montrer leurs conséquences sur le comportement des espèces vis-à-vis du pH du sol. Par contre, nous ne sommes pas parvenus à mettre en évidence une différenciation génétique qui pourrait affecter le comportement d'espèces individuelles dans des régions Vosges et Jura pourtant isolées l'une de l'autre (chapitre 5).

Des phénomènes de compensation de facteurs deviennent facilement observables lorsque l'on étudie le comportement des espèces par rapport à plusieurs facteurs simultanément : la baisse attendue de probabilité de présence d'une espèce lorsque les valeurs d'une variable deviennent moins favorables peut être "compensée" sur un territoire donné grâce aux valeurs d'une seconde variable plus favorable à l'espèce. Ce phénomène peut s'observer aisément quand on compare, dans le chapitre 6, l'évolution de la probabilité de présence de l'Erable champêtre par rapport à trois variables écologiques. Ce type de compensation de facteurs qui conduit apparemment à un changement de comportement de l'espèce lorsqu'on ne prend en compte qu'une variable est expliqué facilement par une approche multi-factorielle de la niche écologique des espèces.

### **Prise en compte simultanée des facteurs du climat et du sol pour déterminer la niche écologique des espèces végétales**

L'apport des variables édaphiques a été illustré de manière spectaculaire pour l'Erable champêtre (*Acer campestre* L.), en terme de qualité du modèle écologique, mais aussi en terme de prédiction de la présence/absence (chapitre 6). L'Erable champêtre réagit avant tout en France par rapport aux facteurs nutritionnels et particulièrement par rapport au pH et à la quantité de calcium présente dans le sol. Sur le plan climatique, il préfère les zones favorables à la croissance d'un point de vue thermique et hydrique (préférence pour des valeurs élevées d'évapotranspiration réelle), mais sans excès de précipitations. Le modèle couplant les variables du sol et du climat permet de beaucoup mieux comprendre le comportement écologique de cette espèce qu'avec une analyse menée avec seulement l'un des deux grands types de facteurs.

L'aire de répartition des espèces dont on prédit la distribution avec des variables du climat est plus uniforme que l'aire de répartition des espèces dont on prédit la distribution avec des variables du sol et du climat (chapitre 6). La prédiction de taches de présence ou d'absence grâce aux variables de sol, qui agit comme un filtre à des échelles locales, permet ainsi d'avoir accès non seulement à l'aire des espèces, mais aussi à leurs habitats potentiels. Cette approche plus complète du comportement des espèces par rapport aux différentes dimensions de leur niche, couplée à la spatialisation des modèles établis, devrait permettre de mieux comprendre l'absence des espèces dans certaines parties du territoire et de distinguer les causes écologiques des causes chorologiques de leur distribution.

La modélisation de la distribution de l'Erable champêtre (*Acer campestre* L.) dans les forêts de France, a montré que les facteurs écologiques permettaient d'expliquer 80 % des présences et des absences de l'espèce (chapitre 6, Coudun et al. sous presse). Il serait intéressant de quantifier la proportion des espèces forestières présentant un tel taux d'explication de la distribution par les facteurs écologiques. Au niveau de chaque espèce, il serait également intéressant de quantifier l'importance relative des facteurs écologiques et des facteurs abiotiques (compétition, dispersion, etc.) ou anthropiques (influence humaine, ancienne occupation du sol, etc.) pour expliquer la distribution observée.

La spatialisation d'un modèle prédictif de distribution d'une espèce végétale à partir des facteurs du climat et des ressources du sol n'avait encore jamais été réalisée. Elle a fourni des résultats très intéressants et amélioré sensiblement les prédictions de distribution réalisées avec le climat seulement. Les possibilités offertes par EcoPlant permettront d'appliquer une procédure similaire à d'autres espèces.

La base de données EcoPlant se révèle très bien adaptée à l'étude de la réponse écologique des espèces et à l'exploration de différentes dimensions de leur niche puisqu'elle stocke de nombreux relevés, et que des variables édaphiques et climatiques sont disponibles en complément de relevés floristiques (chapitre 1, [Gégout et al. 2005](#)). La taille de la base de données et les informations qu'elle contient en font ainsi un excellent outil d'étude de l'écologie et de la distribution des espèces, pour l'instant unique en France et en Europe.

## Perspectives

La possibilité de prédire, avec des modèles établis à une échelle nationale, la distribution ou la présence potentielle des espèces à une échelle locale (le massif, la parcelle) doit encore être étudiée, tout comme la qualité des modèles aux différentes échelles spatiales ([Thuiller et al. 2003a](#)). De même, prédire la distribution future des espèces reste très difficile, puisque des données sur la dispersion des graines, la migration et les capacités d'adaptation des espèces ne sont actuellement pas intégrées dans les modèles de dynamique des espèces à long terme alors qu'elles semblent nécessaires pour réaliser des modèles pertinents. Au delà de l'effet des facteurs écologiques, la colonisation par une espèce des habitats devenus favorables peut être expliquée en partie par des phénomènes de dispersion des espèces, car par exemple une espèce à durée de vie courte et graines légères pourra être dispersée plus rapidement et coloniser plus facilement des nouveaux milieux par rapport à une espèce à durée de vie longue et graines lourdes ([Dirnböck et Dulliger 2004](#)). La critique majeure des modèles prédictifs de la distribution future des espèces, basés sur leur comportement climatique, repose à juste titre sur le fait que les modèles construits supposent un pseudo-équilibre entre l'espèce et son environnement, ce qui incite à des doutes sur leur fiabilité pour prévoir l'évolution future de la distribution des espèces ([Woodward et Beerling 1997](#)).

Dans ce contexte, une perspective souhaitable pour la base de données EcoPlant serait d'être mise en relation avec des bases de données sur les traits de vie des espèces (par exemple LEDA, [Knevel et al. 2003](#)), ce qui pourrait permettre de lier le comportement écologique des espèces à leurs caractéristiques de survie et de dispersion. Une meilleure connaissance des phénomènes de colonisation et d'extinction des espèces forestières pourrait en effet permettre de mieux projeter les modèles statiques de distribution dans des scénarios à moyen ou long terme de changements des conditions climatiques ou trophiques.

## Références bibliographiques



- **Akaike, H. (1973)** Information theory as an extension of the maximum likelihood principle. In : Petrov, B. N. et F. Csaki (eds), *Second Symposium on Information Theory*, 267-281. Akademiai Kiad, Budapest.
- **Allen, R. B., R. K. Peet et W. L. Baker (1991)** Gradient analysis of latitudinal variation in Southern Rocky Mountain forests. *Journal of Biogeography*, **18**, 123-139.
- **Araújo M. B. et P. H. Williams (2000)** Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331-345.
- **Araújo M. B., P. H. Williams et R. J. Fuller (2002)** Dynamics of extinction and the selection of nature reserves. *Proceedings of the Royal Society of London*, **B-269**, 1971-1980.
- **Araújo M. B., M. Cabeza, W. Thuiller, L. Hannah et P. H. Williams (2004)** Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology*, **10**, 1618-1626.
- **Araújo, M. B. et R. G. Pearson (2005)** Equilibrium of species' distributions with climate. *Ecography*, **28**, 693-695.
- **Araújo, M. B., R. G. Pearson, W. Thuiller et M. Erhard (2005)** Validation of species-climate impact models under climate change. *Global Change Biology*, **11**, 1504-1513.
- **Araújo M. B., W. Thuiller et R. G. Pearson (sous presse)** Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography*.
- **Austin, M. P. (1971)** Role of regression analysis in plant ecology. *Proceedings of the Ecological Society of Australia*, **6**, 63-75.
- **Austin, M. P. (1972)** Models and analysis of descriptive vegetation data. *Mathematical models in ecology* (ed. by J. N. R. Jeffers). Blackwell Scientific Publications, Oxford.
- **Austin, M. P., R. B. Cunningham et R. B. Good (1983)** Altitudinal distribution in relation to other environmental factors of several Eucalypt species in Southern New South Wales. *Australian Journal of Ecology*, **8**, 169-180.
- **Austin, M. P., R. B. Cunningham et P. M. Fleming (1984)** New approaches to direct gradient analysis using environmental scalars et statistical curve-fitting procedures. *Vegetatio*, **55**, 11-27.
- **Austin, M. P. (1985)** Continuum concept, ordination methods and niche theory. *Annual Review of Ecology and Systematics*, **16**, 39-61.
- **Austin, M. P. (1987)** Models for the analysis of species response to environmental gradients. *Vegetatio*, **69**, 35-45.
- **Austin, M. P. et T. M. Smith (1989)** A new model for the continuum concept. *Vegetatio*, **83**, 35-47.
- **Austin, M. P., A. O. Nicholls, et C. R. Margules (1990)** Measurement of the realised qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs*, **60**, 161-177.
- **Austin, M. P. et M. J. Gaywood (1994)** Current problems of environmental gradients and species response curves in relation to continuum theory. *Journal of Vegetation Science*, **5**, 473-482.
- **Austin, M. P. et J. A. Meyers (1996)** Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *Forest Ecology and Management*, **85**, 95-106.
- **Austin, M. P. et A. O. Nicholls (1997)** To fix or not to fix the species limits, that is the ecological question: response to Jari Oksanen. *Journal of Vegetation Science*, **8**, 743-748.
- **Austin, M. P. (1998)** An ecological perspective on biodiversity investigations: examples from Australian Eucalypt forests. *Annals of the Missouri Botanical Garden*, **85**, 2-17.
- **Austin, M. P., E. M. Cawsey, B. L. Baker, M. M. Yaleloglou, D. J. Grice, S. V. Briggs, S. Barry, M. D. Doherty, J. Gallant et A. Lehmann (2000)** Predicted vegetation cover in the Central Lachlan region. Final report of the Natural Heritage Trust Project AA 1368.97, CSIRO Wildlife and Ecology, Canberra.
- **Austin, M. P. (2002)** Spatial prediction of species distribution: an interface between ecological theory et statistical modelling. *Ecological Modelling*, **157**, 101-118.
- **Austin, M. P. (soumis)** Predicting species distributions: current problems and potential for progress. *Ecological Modelling*.

- **Badeau, V. (1998)** Caractérisation écologique du réseau européen de suivi des dommages forestiers. Bilan des opérations de terrain et premier résultats. *Les Cahiers du Département de la Santé des Forêts*, **5**. Ministère de l'Agriculture et de la Pêche, Direction de l'Espace Rural et de la Forêt, Paris.
- **Bakkenes, J., J. R. M. Alkemade, F. Ihle, R. Leemans et J. B. Latour (2002)** Assessing the effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology*, **8**, 390-407.
- **Bakker, J. P., R. M. Bekker et K. Thompson (2000)** From seed bank database towards a seed database. *Zeitschrift für Ökologie und Naturschutz*, **9**, 61-72.
- **Bayliss, J. L., V. Simonite et S. Thompson (2005)** The use of probabilistic habitat suitability models for biodiversity action planning. *Agriculture, Ecosystems and Environment*, **108**, 228-250.
- **Becker, M. et N. Le Goff (1988)** Diagnostic stationnel et potentiel de production. *Revue Forestière Française*, **15**, 29-43.
- **Bendixen, K. (2001)** Zum Reproduktionssystem des Feldahorns (*Acer campestre* L.) - Blühphänologie und genetische Untersuchungen. Thèse de doctorat, Institut für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen.
- **Benichou, P. et O. Le Breton (1987)** Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *La Météorologie*, **7**, 23-34.
- **Bergmeier, E. (2002)** The vegetation of the high mountains of Crete: a revision and multivariate analysis. *Phytocoenologia*, **32**, 205-249.
- **Bessemoulin, J. (1989)** *Atlas climatique de la France*. Edition réduite. Direction de la Météorologie Nationale, Paris.
- **Bio, A. M. F., R. Alkemade et A. Barendregt (1998)** Determining alternative models for vegetation response analysis: a non parametric approach. *Journal of Vegetation Science*, **9**, 5-16.
- **Bio, A. M. F. (2000)** Does vegetation suit our models? Data an model assumptions and the assessment of species distribution in space. Thèse de doctorat, University of Utrecht, Faculty of Geographical Science, 207 pp.
- **Bio, A. M. F., P. De Becker, E. De Bie, W. Huybrechts et M. J. Wassen (2002)** Prediction of plant species distribution in lowland river valleys in Belgium: modelling species response to site conditions. *Biodiversity and Conservation*, **11**, 2189-2216.
- **Bonnier, G. (1990)** La grande flore en couleurs. Belin, Paris.
- **Box, E. O., D. W. Crumpacker et E. D. Hardin (1993)** A climatic model for location of plant species in Florida, U.S.A. *Journal of Biogeography*, **20**, 629-644.
- **Box, E. O. (1996)** Plant functional types and climate at the global scale. *Journal of Vegetation Science*, **7**, 309-320.
- **Bragazza, L. et R. Gerdol (1996)** Response surfaces of plant species along water-table depth and pH gradients in a poor mire on the southern Alps (Italy). *Annales Botanici Fennici*, **33**, 11-20.
- **Braun-Blanquet, J. (1932)** Plant sociology, the study of plant communities. McGraw Hill, New-York, NY, US.
- **Brêthes, A. (1989)** La typologie des stations forestières : recommandations méthodologiques. *Revue Forestière Française*, **16**, 7-26.
- **Brisse, H. et G. Grandjouan (1980)** Plantes indicatrices du climat. I. Méthode de caractérisation climatique des plantes à différents niveaux d'abondance. *Bulletin de la Société Botanique de France*, **127**, 263-273.
- **Brisse, H. et M. Kerguélen (1994)** Code Informatisé de la Flore de France (CIFF). Association d'Informatique Appliquée à la Botanique.
- **Brisse, H., P. De Ruffray, G. Grandjouan et M. Hoff (1995)** European Vegetation Survey. The phytosociological database "SOPHY". Part 1. Calibration of indicator plants. Part 2. Socio-ecological classification of the relevés. *Annali Botanica (Roma)*, **53**, 177-223.
- **Brotos, L., W. Thuiller, M. B. Araújo et A. H. Hirzel (2004)** Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437-448.
- **Brzeziecki, B., F. Kienast et O. Wildi (1994)** Potential impacts of a changing climate on the vegetation cover of Switzerland: a simulation experiment using GIS technology. Mountain environments and geographic information

- systems. M. F. Price et D. I. Heywood (eds.), Ch. 14, 263-279.
- **Bugmann, H. K. M. (1994)** On the ecology of mountainous forests in a changing climate: a simulation study. Thèse de Doctorat, Zurich, Switzerland, Swiss Federal Institute of Technology.
  - **Cabeza, M., M. B. Araújo, R. J. Wilson, C. D. Thomas, M. J. R. Cowley et A. Moilanen (2004)** Combining probabilities of occurrence with spatial reserve design. *Journal of Applied Ecology*, **41**, 252-262.
  - **Carpenter, G., A. N. Gillison et J. Winter (1993)** DOMAIN: a flexible modelling procedure for mapping potential distribution of plants and animals. *Biodiversity and Conservation*, **2**, 667-680.
  - **Cawsey, E. M., M. P. Austin et B. L. Baker (2002)** Regional vegetation mapping in Australia: a case study in the practical use of statistical modelling. *Biodiversity and Conservation*, **11**, 2239-2274.
  - **Chase, J. M. et M. A. Leibold (2003)** Ecological niches: linking classical and contemporary approaches. Chicago, The University of Chicago Press.
  - **Cherrill, A. J., C. McClean, P. Watson, K. Tucker, S. P. Rushton et R. Sanderson (1995)** Predicting the distributions of plant species at the regional scale: a hierarchical matrix model. *Landscape Ecology*, **10**, 197-207.
  - **Chytry, M. et M. Rafajova (2003)** Czech national phytosociological database: basic statistics of the available vegetation-plot data. *Preslia*, **75**, 1-15.
  - **Clark, J. S., E. Macklin et L. Wood (1998)** Stages and spatial scales of recruitment limitation in southern Appalachian forests. *Ecological Monographs*, **68**, 213-235.
  - **Clements, F. E. (1936)** Nature and structure of the climax. *Journal of Ecology*, **24**, 253-284.
  - **Collingham, Y. C., R. A. Wadsworth, B. Huntley et P. E. Hulme (2000)** Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology*, **37**, 13-27.
  - **Corley, M. F. V., A. C. Crundwell, R. Düll, O. Hill et A. J. E. Smith (1981)** Mosses of Europe and the Azores: an annotated list of species, with synonyms from the recent literature. *Journal of Bryology*, **11**, 609-689.
  - **Coudun, Ch. (2001)** Vegetal species sensitive to global change in northeast France. Thèse de Master of Science in Environmental Engineering, Department of Environment and Resources. Lyngby, Denmark, Technical University of Denmark, 67 pp.
  - **Coudun, Ch. et J.-C. Gégout (2005)** Ecological behaviour of herbaceous forest species along the pH gradient: a comparison between oceanic and semi-continental regions in northern France. *Global Ecology & Biogeography*, **14**, 263-270.
  - **Coudun, Ch. et J.-C. Gégout (en revision mineure)** The derivation of species response curves with logistic regression models is sensitive to sampling intensity and curve characteristics. *Ecological Modelling*.
  - **Coudun, Ch., J.-C. Gégout, C. Piedallu et J.-C. Rameau (sous presse)** Soil nutritional factors improve plant species distribution models: an illustration with *Acer campestre* L. in France. *Journal of Biogeography*.
  - **Daget, P. et M. Godron (1982)** Analyse fréquentielle de l'écologie des espèces dans les communautés. Collection d'écologie, **18**, Masson, Paris.
  - **Diekmann, M. (1995)** Use et improvement of Ellenberg's indicator values in deciduous forests of the Boreo-nemoral zone in Sweden. *Ecography*, **18**, 178-189.
  - **Diekmann, M. (1996)** Ecological behaviour of deciduous hardwood trees in Boreo-nemoral Sweden in relation to light et soil conditions. *Forest Ecology and Management*, **86**, 1-14.
  - **Diekmann, M. et C. Dupré (1997)** Acidification and eutrophication of deciduous forests in northwestern Germany demonstrated by indicator species analysis. *Journal of Vegetation Science*, **8**, 855-864.
  - **Diekmann, M. et J. E. Lawesson (1999)** Shifts in ecological behaviour of herbaceous forest species along a transect from northern central to north Europe. *Folia Geobotanica*, **34**, 127-141.
  - **Diekmann, M. (2003)** Species indicator values as an important tool in applied plant ecology: a review. *Basic and Applied Ecology*, **4**, 493-506.
  - **Diniz-Filho, J. A. F., L. M. Bini et B. A. Hawkins (2003)** Spatial autocorrelation and red

- herrings in geographical ecology. *Global Ecology & Biogeography*, **12**, 53-64.
- **Dirnböck, T., S. Dullinger et G. Grabherr (2003)** A regional impact assessment of climate and land use change on alpine vegetation. *Journal of Biogeography*, **30**, 1-17.
  - **Dirnböck, T. et S. Dullinger (2004)** Habitat distribution models, spatial autocorrelation, functional traits and dispersal capacity of alpine plant species. *Journal of Vegetation Science*, **15**(1), 77-84.
  - **Duchaufour, P. (1989)** Pédologie et groupes écologiques. I. Rôle du type d'humus et du pH. *Bulletin d'Ecologie*, **20**, 1-6.
  - **Dupré, C. (2000)** How to determine a regional species pool: a study in two Swedish regions. *Oikos*, **89**, 128-136.
  - **Dupré, C., C. Wessberg et M. Diekmann (2002)** Species richness in deciduous forests: effects of species pools and environmental variables. *Journal of Vegetation Science*, **13**, 505-516.
  - **Dzwonko, Z. (2001)** Assessment of light and soil conditions in ancient et recent woodlands by Ellenberg indicator values. *Journal of Applied Ecology*, **38**, 942-951.
  - **Elith, J. (2000)** Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. Quantitative methods in conservation biology. S. Ferson et M. A. Burgman (eds.). New York, Springer, 39-58.
  - **Elith, J. (2002)** Predicting the distribution of plants. Thèse de doctorat, University of Melbourne, Australia.
  - **Elith, J. et M. A. Burgman (2002)** Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. Predicting species occurrences: issues of accuracy and scale. J. M. Scott, P. J. Heglund, F. Samson J. Haufler, M. Morrison, M. Raphael et B. Wall (eds.). Covelo, California, Island Press: Ch. 24, 303-313.
  - **Elith, J., M. A. Burgman et H. M. Regan (2002)** Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling*, **157**, 313-329.
  - **Ellenberg, H., H. E. Weber, R. Düll, V. Wirth, W. Werner et D. Paulißen (1992)** Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta Geobotanica*, **18**, 1-248.
  - **Eriksson, O. (1993)** The species-pool hypothesis and plant community diversity. *Oikos*, **68**, 371-374.
  - **Ertsen, A. C. D., J. R. M. Alkemade et M. Wassen (1998)** Calibrating Ellenberg indicator values for moisture, acidity, nutrient availability et salinity in the Netherlands. *Plant Ecology*, **135**, 113-124.
  - **Espiau, P. et A. Peyronel (1977)** Acidité d'échange dans les sols. Application à une séquence altitudinale des sols du massif du Mont Aigoual. *Science du Sol*, **4**, 25-44.
  - **ESRI (2002)** ArcGIS 8.2. Environmental Systems Research Institutes, Redlands, California.
  - **Ewald, J. (2001)** Der Beitrag pflanzensoziologischer Datenbanken zur vegetationsökologischen Forschung. *Berichte der Rheinhold-Tüxen Gesellschaft*, **13**, 53-69.
  - **Ewald, J. (2003)** The calcareous riddle: why are there so many calciphilous species in the Central European flora? *Folia Geobotanica*, **38**, 357-366.
  - **Falkengren-Grerup, U., J. Brunet, M. E. Quist et G. Tyler (1995)** Is the Ca:Al ratio superior to pH, Ca or Al concentrations of soils in accounting for the distribution of plants in deciduous forests? *Plant and Soil*, **177**, 21-31.
  - **Ferrier, S., G. Watson, J. Pearce et M. Drielsma (2002)** Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity et Conservation*, **11**, 2275-2307.
  - **Fertig, W. et W. A. Reiners (2002)** Predicting presence/absence of plant species for range mapping: a case study from Wyoming. Predicting species occurrences: Issues of accuracy and scale. J. M. Scott, P. J. Heglund, F. Samson J. Haufler, M. Morrison, M. Raphael et B. Wall (eds.). Covelo, California, Island Press: Ch. 42, 483-489.
  - **Fielding, A. H. et J. F. Bell (1997)** A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38-49.
  - **Firbank, L. G., N. E. Ellis, M. O. Hill, A. J. Lockwood et R. D. Swetnam (1998)** Mapping the distribution of weeds in Great Britain in relation to national survey data and to soil type. *Weed Research*, **38**, 1-10.

- **Fortin, M.-J., T. H. Keitt, B. A. Maurer, M. L. Taper, D. M. Kaufman et T. M. Blackburn (2005)** Species' geographic ranges and distributional limits: pattern analysis and statistical issues. *Oikos*, **108**, 7-17.
- **Franklin, J. (1995)** Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, 474-499.
- **Franklin, J. (1998)** Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *Journal of Vegetation Science*, **9**, 733-748.
- **Franklin, J., P. McCullough et C. Gray (2000)** Terrain variables used for predictive mapping of vegetation communities in Southern California. Terrain analysis: principles and applications. J. P. Wilson and J. C. Gallant (eds.). New York, Wiley.
- **Franklin, J. (2002)** Enhancing a regional vegetation map with predictive models of dominant plant species in chaparral. *Applied Vegetation Science*, **5**, 135-146.
- **Garbolino, E. (2001)** Les plantes indicatrices du climat en France et leur télédétection. Thèse de doctorat. Université de Nice-Sophia Antipolis, 310 pp.
- **Gégout, J.-C. (1995)** Etude des relations entre les ressources minérales du sol et la végétation forestière dans les Vosges. Thèse de doctorat, Université de Nancy I, 215 pp.
- **Gégout, J.-C. et J.-C. Pierrat (1998)** L'autécologie des espèces végétales : une approche par régression non paramétrique. *Ecologie*, **29**, 473-482.
- **Gégout, J.-C. (2001)** Création d'une base de données phytoécologiques pour déterminer l'autécologie des espèces de la flore forestière de France. *Revue Forestière Française*, **53**, 397-403.
- **Gégout, J.-C. et B. Jabiol (2001)** Analyses de sols en forêt : les choix du phytoécologue dans le cadre des typologies de stations ou des études scientifiques. *Revue Forestière Française*, **53**, 568-580.
- **Gégout, J.-C., Ch. Coudun, H. Brisse et L. Bergès (2002)** Comportement écologique des espèces forestières vis-à-vis du climat et du sol en France: Application à l'évaluation des charges critiques d'acidité et d'azote. Rapport remis à l'ADEME, ENGREF, Nancy, 51 pp.
- **Gégout, J.-C. et E. Krizova (2003)** Comparison of indicator values of forest understory plant species in Western Carpathians (Slovakia) and Vosges Mountains (France). *Forest Ecology and Management*, **182**, 1-11.
- **Gégout, J.-C., J.-C. Hervé, F. Houllier et J.-C. Pierrat (2003)** Prediction of forest soil nutrient status using vegetation. *Journal of Vegetation Science*, **14**, 55-62.
- **Gégout, J.-C., Ch. Coudun, G. Bailly et B. Jabiol (2005)** EcoPlant: a forest sites database to link floristic data with soil resources and climatic conditions. *Journal of Vegetation Science*, **16**, 257-260.
- **Gelfand, A. E., J. A. Silander Jr., S. Wu, A. Latimer, P. O. Lewis, A. G. Rebelo et M. Holder (2003)** Explaining species distribution patterns through hierarchical modelling. *Bayesian Analysis*, **1**, 1-47.
- **Gignac, L. D., D. H. Vitt et S. E. Bayley (1991)** Bryophyte response surfaces along ecologic and climatic gradients. *Vegetatio*, **93**, 29-45.
- **Gignac, L. D. (1992)** Niche structure, resource partitioning and species interactions of mire bryophytes relative to climatic and ecological gradients in western Canada. *The Bryologist*, **95**, 406-418.
- **Gleason, H. A. (1917)** The structure and development of the plant association. *Bulletin of the Torrey Botanical Club*, **44**, 463-481.
- **Gleason, H. A. (1926)** The individualistic concept of plant association. *Bulletin of the Torrey Botanical Club*, **53**, 7-26.
- **Goodall, D. W. (1963)** The continuum and the individualistic association. *Vegetatio*, **11**, 297-316.
- **Gough, L., G. R. Shaver, J. Carroll, D. L. Royer et J. A. Laundre (2000)** Vascular plant species richness in Alaskan arctic tundra: the importance of soil pH. *Journal of Ecology*, **88**, 54-66.
- **Gower, J. C. et P. Legendre (1986)** Metric and euclidean properties of dissimilarity coefficients. *Journal of Classification*, **3**, 5-48.
- **Grace, J. B. (1999)** The factors controlling species density in herbaceous plant communities:

- an assessment. *Perspectives in Plant Ecology, Evolution and Systematics*, **2**, 1-28.
- **Grace, J. B. (2001)** Difficulties with estimating and interpreting species pools and the implications for understanding patterns of diversity. *Folia Geobotanica*, **36**, 71-83.
  - **Grime, J. P. (1979)** Plant strategies and vegetation processes. London, Wiley.
  - **Grolle, R. (1983)** Hepatics of Europe including the Azores. An annotated list of species, with synonyms from the recent literature. *Journal of Bryology*, **12**, 403-459.
  - **Guisan, A., J.-P. Theurillat et F. Kienast (1998)** Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science*, **9**, 65-74.
  - **Guisan, A., S. B. Weiss and A. D. Weiss (1999)** GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, **143**, 107-122.
  - **Guisan, A. et J.-P. Theurillat (2000a)** Assessing alpine plant vulnerability to climate change: a modeling perspective. *Integrated Assessment*, **1**, 307-320.
  - **Guisan, A. et J.-P. Theurillat (2000b)** Equilibrium modeling of alpine plant distribution and climate change: how far can we go? *Phytocoenologia*, **30**, 353-384.
  - **Guisan, A. et N. E. Zimmermann (2000)** Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.
  - **Guisan, A., T. C. Edwards Jr. et T. J. Hastie (2002)** Generalized linear et generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89-100.
  - **Guries, R. P. et E. V. Nordheim (1984)** Flight characteristics and dispersal potential of Maple samaras. *Forest Science*, **30**, 434-440.
  - **Hampe, A. (2004)** Bioclimate envelope models: what they detect and what they hide. *Global Ecology & Biogeography*, **13**, 469-471.
  - **Hastie, T. J. et R. Tibshirani (1997)** Generalized additive models. London, Chapman and Hall.
  - **Hawkes, J. C., D. G. Pyatt et I. M. S. White (1997)** Using Ellenberg indicator values to assess soil quality in British forests from ground vegetation: A pilot study. *Journal of Applied Ecology*, **34**, 375-387.
  - **Heegaard, E. (1997)** Ecology of *Andreaea* in western Norway. *Journal of Bryology*, **19**, 527-636.
  - **Heegaard, E. (2001)** Environmental relationships of perichaetial and sporophyte production in *Andreaea* spp. in Western Norway. *Journal of Bryology*, **23**, 97-108.
  - **Heegaard, E., H. J. B. Birks, C. E. Gibson, S. J. Smith et S. Wolfe-Murphy (2001)** Species-environmental relationships of aquatic macrophytes in Northern Ireland. *Aquatic Botany*, **70**, 175-223.
  - **Heegaard, E. (2002a)** The outer border and central border for species-environmental relationships estimated by non-parametric generalised additive models. *Ecological Modelling*, **157**, 131-139.
  - **Heegaard, E. (2002b)** A model of alpine species distribution in relation to snowmelt time and altitude. *Journal of Vegetation Science*, **13**, 493-504.
  - **Hennekens, S. et J. H. J. Schaminée (2001)** TURBOVEG, a comprehensive data base management system for vegetation data. *Journal of Vegetation Science*, **12**, 589-591.
  - **Hill, M. O. et P. D. Carey (1997)** Prediction of yield in the Rothamsted Park Grass Experiment by Ellenberg indicator values. *Journal of Vegetation Science*, **8**, 579-586.
  - **Hill, M. O., J. O. Mountford, D. B. Roy et R. G. H. Bunce (1999)** Ellenberg's indicator values for British plants. Technical Annex to Volume 2 of the ECOFACT research report series., CEH - Center of Ecology and Hydrology, Natural Environment Research Council, 46 pp.
  - **Hill, M. O., D. B. Roy, J. O. Mountford et R. G. H. Bunce (2000)** Extending Ellenberg's indicator values to a new area: an algorithmic approach. *Journal of Applied Ecology*, **37**, 3-15.
  - **Hooten, M. B., D. R. Larsen et C. K. Wikle (2003)** Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology*, **18**, 487-502.
  - **Huisman, J., H. Olf et L. F. M. Fresco (1993)** A hierarchical set of models for species response analysis. *Journal of Vegetation Science*, **4**, 37-46.

- **Huntley, B., R. E. Green, Y. C. Collingham, J. K. Hill, S. G. Willis, P. J. Bartlein, W. Cramer, W. J. Hagemeyer et C. J. Thomas (2004)** The performance of models relating species geographical distributions to climate is independent of trophic level. *Ecology Letters*, **7**, 417-426.
- **Hutchinson, G. E. (1957)** Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology.
- **Isaaks, E. H. et R. M. Srivastava (1989)** *An introduction to applied geostatistics*. Oxford University Press, New York.
- **Iverson, L. R., A. Prasad et M. W. Schwartz (1999)** Modeling potential future individual tree-species distributions in the eastern United States under a climate change scenario: a case study with *Pinus virginiana*. *Ecological Modelling*, **115**, 77-93.
- **Jäger, E. J. (2000)** A database on biological traits of the German flora: state of the art and need of investigation of the vegetative structures. *Zeitschrift für Ökologie und Naturschutz*, **9**, 53-59.
- **Jamagne, M., R. Bétrémieux, J. C. Bégon et A. Mori (1977)** Quelques données sur la variabilité dans le milieu naturel de la réserve en eau des sols. *B.T.I. du Ministère de l'Agriculture*, **324-325**, 627-641.
- **Journel, A. G. (1983)** Nonparametric estimation of spatial distributions. *Mathematical Geology*, **15**, 445-468.
- **Karl, J. W., L. K. Svancara, P. J. Heglund, N. M. Wright and J. M. Scott (2002)** Species commonness and the accuracy of habitat-relationship models. Predicting species occurrences: Issues of accuracy and scale. J. M. Scott, P. J. Heglund, F. Samson J. Haufler, M. Morrison, M. Raphael et B. Wall (eds.). Covelo, California, Island Press: Ch. 51, 573-580.
- **Knevel, I. C., R. M. Bekker, J. P. Bakker et M. Kleyer (2003)** Life-history traits of the Northwest European flora: a data-base (LEDA). *Journal of Vegetation Science*, **14**, 611-614.
- **Koenig, W. D. (1999)** Spatial autocorrelation of spatial phenomena. *Trends in Ecology & Evolution*, **14**, 22-26.
- **Körner, C. (1995)** Alpine plant diversity: a global survey and functional interpretations. Arctic and alpine biodiversity: patterns, causes and ecosystem consequences. F. S. Chapin III. and C. Körner (eds.). Berlin, Springer. 113: 45-62.
- **Landolt, E. (1977)** Ökologische Zeigerwerte zur Schweizer Flora. Zürich. Veröffentlichungen des geobotanischen Institutes der ETH - Stiftung Rübél ; Heft 64. 208 pp.
- **Lawesson, J. E., M. Diekmann, O. Eilertsen, A. M. Fosaa et H. Heikkilä (1997)** The Nordic Vegetation Survey: concepts and perspectives. *Journal of Vegetation Science*, **8**, 455-458.
- **Lawesson, J. E. et S. Mark (2000)** pH and Ellenberg reaction values for Danish forest plants. *Proceedings LAVS Symposium*, 151-153.
- **Lawesson, J. E. et J. Oksanen (2002)** Niche characteristics of Danish woody species as derived from coenoclines. *Journal of Vegetation Science*, **13**, 279-290.
- **Le Duc, M. G., M. O. Hill et T. H. Sparks (1992)** A method for predicting the probability of species occurrence using data from systematic surveys. *Watsonia*, **19**, 97-105.
- **Le Tacon, F. et J. Timbal (1972)** A propos des conditions écologiques des hêtraies dans le nord-est et le nord-ouest de la France. *Revue Forestière Française*, **24**, 187-200.
- **Le Tacon, F. et J. Timbal (1973)** Valeurs indicatrices des principales espèces végétales des hêtraies du Nord-Est de la France, vis-à-vis des types d'humus. *Revue Forestière Française*, **25**(4), 269-281.
- **Leathwick, J. R. et G. M. Rogers (1996)** Modelling relationships between environment and canopy composition in secondary vegetation in Central North Island, New Zealand. *New Zealand Journal of Ecology*, **20**, 147-161.
- **Leathwick, J. R. (1998)** Are New-Zealand's Nothofagus species in equilibrium with their environment? *Journal of Vegetation Science*, **9**, 719-732.
- **Leathwick, J. R. et D. Whitehead (2001)** Soil and atmospheric water deficits and the distribution of New Zealand's indigenous tree species. *Functional Ecology*, **15**, 233-242.
- **Legendre, P. (1993)** Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659-1673.
- **Lehmann, A., J. McC. Overton et M. P. Austin (2002a)** Regression models for spatial prediction: their role for biodiversity and

- conservation. *Biodiversity and Conservation*, **11**, 2085-2092.
- **Lehmann, A., J. R. Leathwick and J. McC. Overton (2002b)** Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodiversity and Conservation*, **11**, 2217-2238.
  - **Lehmann, A., J. McC. Overton et J. R. Leathwick (2002c)** GRASP: Generalized regression analysis and spatial prediction. *Ecological Modelling*, **157**, 189-207.
  - **Lenihan, J. M. (1993)** Ecological response surfaces for North American boreal tree species and their use in forest classification. *Journal of Vegetation Science*, **4**, 667-680.
  - **Lennon, J. J. (2000)** Red-shifts and red herrings in geographical ecology. *Ecography*, **23**, 101-113.
  - **Lipsett-Moore, G., D. W. McKenney and S. Jones (2003)** Multi-scale species modelling in Ontario: a workshop on needs and opportunities. *The Forestry Chronicle*, **79**, 147-148.
  - **Luoto, L., R. K. Heikkinen, J. Pöyry et K. Saarinen (sous presse)** Determinants of biogeographical distribution of butterflies in boreal regions. *Journal of Biogeography*.
  - **Maggini, R., A. Lehmann, N. E. Zimmermann et A. Guisan (sous presse)** Improving generalized regression analysis for spatial predictions of forest communities. *Journal of Biogeography*.
  - **Malanson, G. P. (1997)** Simulated responses to hypothetical fundamental niches. *Journal of Vegetation Science*, **8**, 307-316.
  - **Manel, S., J.-M. Dias, S. T. Buckton et S. J. Ormerod (1999a)** Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, **36**, 734-747.
  - **Manel, S., J.-M. Dias et S. J. Ormerod (1999)** Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, **120**, 337-347.
  - **Manel, S., H. C. Williams et S. J. Ormerod (2001)** Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921-931.
  - **Margules, C. R. et J. L. Stein (1989)** Patterns in the distributions of species and selection of nature reserves: an example from Eucalyptus forests in southeastern New South Wales. *Biological Conservation*, **50**, 219-238.
  - **Marinoni, O. (2003)** Improving geological models using a combined ordinary-indicator kriging approach. *Engineering Geology*, **69**, 37-45.
  - **MathSoft, Inc. (1999)** S-Plus 2000, Programmer's Guide. Seattle, Washington, United States of America, MathSoft, Inc.
  - **McCullagh, P. et J. A. Nelder (1997)** Generalized linear models. 2<sup>nd</sup> edn, Chapman & Hall, London.
  - **McKenney, D. W., L. A. Venier, A. Heerdegen et M. A. McCarthy (2002)** A Monte Carlo experiment for species mapping problems. Predicting species occurrences: Issues of accuracy and scale. J. M. Scott, P. J. Heglund, F. Samson J. Haufler, M. Morrison, M. Raphael et B. Wall (eds.). Covelo, California, Island Press: Ch. 31, 377-381.
  - **McKenzie, D. et C. B. Halpern (1999)** Modeling the distributions of shrub species in Pacific northwest forests. *Forest Ecology and Management*, **114**, 293-307.
  - **McKenzie, D., D. W. Peterson et D. L. Peterson (2003a)** Modelling conifer species distributions in mountain forests of Washington State, USA. *The Forestry Chronicle*, **79**, 253-258.
  - **McKenzie, D., D. W. Peterson, D. L. Peterson et P. E. Thornton (2003b)** Climatic and biophysical controls on conifer species distributions in mountain forests of Washington State, USA. *Journal of Biogeography*, **30**, 1093-1108.
  - **McPherson, J. M., W. Jetz and D. J. Rogers (2004)** The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811-823.
  - **Miller, J. et J. Franklin (2002)** Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*, **157**, 227-247.
  - **Mills, E. (1996)** An appreciation and natural history of the English field maple (*Acer campestre* L.). *Arboricultural Journal*, **20**, 405-410.

- **Mladenoff, D. J., T. A. Sickley et A. P. Wydeven (1999)** Predicting gray wolf landscape recolonization: Logistic regression models vs. new field data. *Ecological Applications*, **9**, 37-44.
- **Mohler, C. L. (1983)** Effect of sampling pattern on estimation of species distributions along gradients. *Vegetatio*, **54**, 97-102.
- **Motel, G. (1995)** L'érable. Actes Sud, Paris.
- **Mucina, L., J. S. Rodwell, J. H. J. Schaminée et H. Dierschke (1993)** European Vegetation Survey: Current state of some national programmes. *Journal of Vegetation Science*, **4**, 429-438.
- **Mucina, L., G. J. Bredenkamp, D. B. Hoare et D. J. McDonald (2000a)** A national vegetation database for South Africa. *South African Journal of Science*, **96**, 497-498.
- **Mucina, L., J. H. J. Schaminée et J. S. Rodwell (2000b)** Common data standards for recording relevés in field survey for vegetation classification. *Journal of Vegetation Science*, **11**, 769-772.
- **Munoz, J. et A. M. Felicísimo (2004)** Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science*, **15**, 285-292.
- **Noirfalise, A. et M. Dethioux (1970)** Répertoire écologique des espèces forestières de Belgique. Gembloux, Centre d'Ecologie Forestière, 28 pp.
- **Odland, A., H. J. B. Birks et J. M. Line (1995)** Ecological optima et tolerances of *Thelypteris limbosperma*, *Athyrium distentifolium*, et *Mattuccia struthiopteris* along environmental gradients in Western Norway. *Vegetatio*, **120**, 115-129.
- **Oksanen, J. (1997)** Why the beta-function cannot be used to estimate skewness of species responses. *Journal of Vegetation Science*, **8**, 147-152.
- **Oksanen, J., E. Läärä, K. Tolonen et B. G. Warner (2001)** Confidence intervals for the optimum in the Gaussian response function. *Ecology*, **82**, 1191-1197.
- **Parrish, J. A. D. et F. A. Bazzaz (1985)** Ontogenetic niche shifts in old-field annuals. *Ecology*, **66**, 1296-1302.
- **Pärtel, M., M. Zobel, K. Zobel et E. Van der Maarel (1996)** The species pool and its relation to species richness: evidence from Estonian plant communities. *Oikos*, **75**, 111-117.
- **Pärtel, M. (2002)** Local plant diversity patterns and evolutionary history at the regional scale. *Ecology*, **83**, 2361-2366.
- **Pearce, J. et S. Ferrier (2000a)** An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127-147.
- **Pearce, J. et S. Ferrier (2000b)** Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225-245.
- **Pearson, R. G., T. P. Dawson, P. M. Berry et P. A. Harrison (2002)** SPECIES: a spatial evaluation of climate impact on the envelope of species. *Ecological Modelling*, **154**, 289-300.
- **Peet, R. K., J. D. Fridley et J. M. Gramling (2003)** Variation in species richness and species pool size across a pH gradient in forests of the southern Blue Ridge Mountains. *Folia Geobotanica*, **38**, 391-401.
- **Peet, R. K. et S. K. Wiser (2004)** Databases and information systems for vegetation science. Proceedings of the 2004 Symposium, IAVS Working Group for Ecoinformatics (<http://www.bio.unc.edu/faculty/peet/vegdata/iavs2004.htm>).
- **Penel, M. (1979)** Caractérisation physico-chimique et classification des humus forestiers acides en relation avec la végétation et ses exigences écologiques.
- **Peterson, A. T. (2003)** Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology*, **78**, 419-433.
- **Peterson, A. T., M. Papes et D. A. Kluza (2003)** Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science*, **51**, 863-868.
- **Philip, G. M. et D. F. Watson D. F. (1982)** A precise method for determining contoured surfaces. *Australian Petroleum Exploration Association Journal*, **22**, 205-212.
- **Pinto, P. et J.-C. Gégout (2005)** Assessing the nutritional and climatic response of temperate forest tree species in the Vosges Mountains. *Annals of Forest Science*, **62**, 1-10.

- Polasky, S., J. D. Camm, A. R. Solow, B. Csuti, D. White et R. Ding (2000) Choosing reserve networks with incomplete species information. *Biological Conservation*, **94**, 1-10.
- Plantureux, S. (1996) FLORA-sys : système informatique de gestion et d'aide à l'interprétation des relevés floristiques. *Acta Botanica Gallica*, **143**, 403-410.
- Poschlod, P., M. Kleyer, A.-K. Jackel, A. Dannemann, et O. Tackenberg (2003) BIOPOP - a database of plant traits and Internet application for nature conservation. *Folia Geobotanica*, **38**, 263-271.
- Prinzing, A., W. Durka, S. Klotz et R. Brandl (2002) Geographic variability of ecological niches of plant species: are competition et stress relevant? *Ecography*, **25**, 721-729.
- Py, N. (2005) Cartographie prédictive de l'aire de répartition et de la fertilité de quelques essences forestières. Application à l'érable champêtre (*Acer campestre*), au hêtre (*Fagus sylvatica*) et à l'épicéa (*Picea abies*) dans les Alpes françaises. Ecole Nationale du Génie Rural, des Eaux et des Forêts (ENGREF), Nancy, France.
- Rameau, J.-C., D. Mansion, G. Dumé, J. Timbal, A. Lecoite, P. Dupont et R. Keller (1989) Flore forestière française. Guide écologique illustré. Vol 1, plaines et collines. Institut pour le Développement Forestier, Paris.
- Rameau, J.-C., D. Mansion, G. Dumé A. Lecoite, J. Timbal, P. Dupont et R. Keller (1993) Flore forestière française. Guide écologique illustré. Vol 2, montagnes. Institut pour le Développement Forestier, Paris.
- Randin, C. F., T. Dirnböck, S. Dullinger, N. E. Zimmermann, M. Zappa et A. Guisan (sous presse) Are niche distribution models transferable in space? *Journal of Biogeography*.
- Robertson, M. P., M. H. Villet et A. R. Palmer (2004) A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous insects. *Diversity and Distributions*, **10**, 461-474.
- Roy, D. B., M. O. Hill, P. Rothery et R. G. H. Bunce (2000) Ecological indicator values of British species: an application of Gaussian logistic regression. *Annales Botanici Fennici*, **37**, 219-226.
- Rushton, S. P., S. J. Ormerod et G. Kerby (2004) New paradigms for modelling species distributions? *Journal of Applied Ecology*, **41**, 193-200.
- Rydgren, K., R. H. Økland et T. Økland (2003) Species response curves along environmental gradients. A case study from SE Norwegian swamp forests. *Journal of Vegetation Science*, **14**, 869-880.
- Schaffers, A. P. et K. V. Sykora (2000) Reliability of Ellenberg indicator values for moisture, nitrogen et soil reaction: A comparison with field measurements. *Journal of Vegetation Science*, **11**, 225-244.
- Schröder, B. et O. Richter (1999) Are habitat models transferable in space and time? *Zeitschrift für Ökologie und Naturschutz*, **8**, 195-205.
- Schuster, B. et M. Diekmann (2003) Changes in species density along the soil pH gradient - evidence from German plant communities. *Folia Geobotanica*, **38**, 367-379.
- Schuster, B. et M. Diekmann (2005) Species richness and environmental correlates in deciduous forests of Northwest Germany. *Forest Ecology and Management*, **206**, 197-205.
- Schwartz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464.
- Schwartz, M. W., L. R. Iverson et A. M. Prasad (2001) Predicting the potential future distribution of four tree species in Ohio using current habitat availability and climatic forcing. *Ecosystems*, **4**, 568-581.
- Schwarz, P. A., T. J. Fahey et C. E. McCulloch (2003) Factors controlling spatial variation of tree species abundance in a forested landscape. *Ecology*, **84**, 1862-1878.
- Scott, J. M., P. J. Heglund, F. Samson, J. Haufler, M. Morrison, M. Raphael et B. Wall (2002) Predicting species occurrences: Issues of accuracy and scale. Covelo, California, Island Press.
- Segurado, P. et M. B. Araújo (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**, 1-14.
- Segurado, P., M. B. Araújo et W. E. Kunin (soumis) Consequences of spatial autocorrelation on niche-based models. *Journal of Applied Ecology*.
- Shao, G. et P. N. Halpin (1995) Climatic controls of eastern North American coastal tree

- and shrub distributions. *Journal of Biogeography*, **22**, 1083-1089.
- **Skov, F. et J.-C. Svenning (2004)** Potential impact of climatic change on the distribution of forest herbs in Europe. *Ecography*, **27**, 366-380.
  - **Stockwell, D. R. B. and A. T. Peterson (2002a)** Controlling bias in biodiversity data. Predicting species occurrences: Issues of accuracy and scale. J. M. Scott, P. J. Heglund, F. Samson J. Haufler, M. Morrison, M. Raphael et B. Wall (eds.). Covelo, California, Island Press: Ch. 48, 537-546.
  - **Stockwell, D. R. B. et A. T. Peterson (2002b)** Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1-13.
  - **Swets, J. A. (1988)** Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285-1293.
  - **ter Braak, C. J. F. et L. G. Barendregt (1986)** Weighted averaging of species indicator values: its efficiency in environmental calibration. *Mathematical Biosciences*, **78**, 57-72.
  - **ter Braak, C. J. F. et C. W. N. Looman (1986)** Weighted averaging, logistic regression et the Gaussian response model. *Vegetatio*, **65**, 3-11.
  - **ter Braak, C. J. F. et N. J. M. Gremmen (1987)** Ecological amplitudes of plant species et the internal consistency of Ellenberg's indicator values for moisture. *Vegetatio*, **69**, 79-87.
  - **ter Braak, C. J. F. (1996)** Unimodal models to relate species to environment. Agricultural Mathematics Group.
  - **Theurillat, J.-P. et A. Guisan (2001)** Potential impact of climate change on vegetation in the European Alps: a review. *Climatic Change*, **50**, 77-109.
  - **Thimonier, A., J.-L. Dupouey, F. Bost et M. Becker (1994)** Simultaneous eutrophication and acidification of a forest ecosystem in North-East France. *New Phytologist*, **126**, 533-539.
  - **Thompson, K., J. G. Hodgson, J. P. Grime, I. H. Rorison, S. R. Band et R. E. Spencer (1993)** Ellenberg numbers revisited. *Phytocoenologia*, **23**, 277-289.
  - **Thornthwaite, C. W. et J. R. Mather (1957)** Instructions and tables for computing potential evapotranspiration and the water balance. *Publications in Climatology*, **10**, 183-311.
  - **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., M. B. Araújo et S. Lavorel (2003a)** Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, **14**, 669-680.
  - **Thuiller, W., J. Vayreda, J. Pino, S. Sabate, S. Lavorel and C. Gracia (2003b)** Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology & Biogeography*, **12**, 313-325.
  - **Thuiller, W., L. Brotons, M. B. Araújo et S. Lavorel (2004)** Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165-172.
  - **Tichy, L. (2002)** JUICE, software for vegetation classification. *Journal of Vegetation Science*, **13**, 451-453.
  - **Turc, L. (1961)** Evaluation des besoins en eau d'irrigation et évaporation potentielle. *Annales Agronomiques*, **12**, 13-49.
  - **Turesson, G. (1922)** The species et the variety as ecological units. *Hereditas*, **3**, 100-113.
  - **Tutin, T. G., V. H. Heywood, N. A. Burges, D. H. Valentine, S. M. Walters et D. A. Webb (2001)** Flora Europaea, vols. 1-5. Cambridge University Press, Cambridge, UK.
  - **Tyler, G. (1999)** Plant distribution and soil-plant interactions on shallow soils. Acta Phytogeographica Suecica. H. Rydin, P. Snoeijis et M. Diekmann (eds.). 84: Ch. 3, 21-32.
  - **Tyler, G. (2003)** Some ecophysiological et historical approaches to species richness et calcicole/calcifuge behaviour – contribution to a debate. *Folia Geobotanica*, **38**, 419-428.
  - **Ulrich, E. (1997)** Organization of forest system monitoring in France - the RENECOFOR network. World Forestry Congress, Antalya, Turkey.
  - **Vanmechelen, L., R. Groenemans et E. Van Ranst (1997)** Forest soil condition in Europe. Results of a large-scale soil survey. Forest Soil Co-Ordinating Centre.

- **Vayssières, M. P., R. E. Plant et B. H. Allen-Diaz (2000)** Classification trees: An alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, **11**, 679-694.
- **Vetaas, O. R. (2000)** Comparing species temperature response curves: population density versus second-hand data. *Journal of Vegetation Science*, **11**, 659-666.
- **Vetaas, O. R. (2002)** Realized and potential climate niches: a comparison of four Rhododendron tree species. *Journal of Biogeography*, **29**, 545-554.
- **Virtanen, A., V. Kairisto et E. Uusipaikka (1998)** Regression-based reference limits: determination of sufficient sample size. *Clinical Chemistry*, **44**, 2353-2358.
- **Walter, H. et E. Walter (1953)** Einige allgemeine Ergebnisse unserer Forschungsreise nach Südwestafrika 1952/1953: das Gesetz der relativen Standortskonstanz; das Wesen der Pflanzengemeinschaften. *Ber. Deutsch. Bot. Ges.*, **66**, 227-235.
- **Wamelink, G. W. W., V. Joosten, H. F. Van Dobben et F. Berendse (2002)** Validity of Ellenberg indicator values judged from physico-chemical field measurements. *Journal of Vegetation Science*, **13**, 269-278.
- **Werger, M. J. A. et E. Van der Maarel (1978)** Plant species and plant communities: some conclusions. Plant species and plant communities. E. Van der Maarel et M. J. A. Werger (eds.). The Hague, Junk.
- **Whittaker, R. H. (1967)** Gradient analysis of vegetation. *Biological Reviews*, **49**, 207-264.
- **Williams, P. H., L. Hannah, S. Andelman, G. F. Midgley, M. B. Araújo, G. Hughes, L. L. Manne, E. Martinez-Meyer et R. G. Pearson (2005)** Planning for climate change: identifying minimum-dispersal corridors for the Cape Proteaceae. *Conservation Biology*, **19**, 1063-1074.
- **Wiser, S., R. K. Peet et P. S. White (1998)** Prediction of rare-plant occurrence: a southern Appalachian example. *Ecological Applications*, **8**, 909-920.
- **Wiser, S. K., P. J. Bellingham et L. E. Burrows (2001)** Managing biodiversity information: Development of New Zealand's National Vegetation Survey databank. *New Zealand Journal of Ecology*, **25**, 1-17.
- **Woodward, F. I. et B. G. Williams (1987)** Climate and plant distribution at global and local scales. *Vegetatio*, **69**, 189-197.
- **Woodward, F. I. et D. J. Beerling (1997)** The dynamics of vegetation change: health warnings for equilibrium "dodo" models. *Global Ecology and Biogeography Letters*, **6**, 413-418.
- **Yee, T. W. et N. D. Mitchell (1991)** Generalized additive models in plant ecology. *Journal of Vegetation Science*, **2**, 587-602.
- **Zaniewski, A. E., A. Lehmann et J. McC. Overton (2002)** Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, **157**, 261-280.
- **Zimmermann, N. E. et F. Kienast (1999)** Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science*, **10**, 469-482.
- **Zimmermann, N. E., G. G. Moisen, T. C. Edwards Jr., T. S. Frescino et J. A. Blackard (soumis)** Testing the partial contributions of remotely sensed and topoclimatic predictors for tree species habitat modelling in Utah. *Journal of Applied Ecology*.
- **Zobel, M. (1997)** The relative role of species pools in determining plant species richness: an alternative explanation of species coexistence? *Trends in Ecology and Evolution*, **12**, 266-269.
- **Zobel, M., E. Van der Maarel et C. Dupré (1998)** Species pool: the concept, its first determination and significance for community restoration. *Applied Vegetation Science*, **1**, 55-66.