

N° attribué par la bibliothèque : |_|_|_|_|_|_|_|_|_|_|

**UNIVERSITE MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC**

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline : Biologie des Populations et Ecologie

Formation Doctorale : Biologie de l'Evolution et Ecologie

Ecole Doctorale : Biologie Intégrative

Impact des changements globaux sur la biodiversité en Europe : projections et incertitudes

Par

Wilfried THUILLER

Soutenue publiquement le 15 décembre 2003, devant le jury composé de :

M. Antoine GUISAN, Professeur assistant, Université de Lausanne	Rapporteur
M. Jaume TERRADAS, Professeur, Université Autonome de Barcelone	Rapporteur
M. Philippe CHOLER, Maître de Conférence, Université J. Fourier Grenoble	Examineur
M. Joël GUIOT, Directeur de Recherche, CEREGE Aix en Provence	Examineur
M. Doyle MCKEY, Professeur, Université Montpellier II	Examineur
Mme Sandra LAVOREL, Directeur de Recherche, Université J. Fourier Grenoble	Directeur de thèse

REMERCIEMENTS

Cette thèse a été le travail de nombreuses collaborations, de découvertes et de rencontres, il me sera donc impossible de remercier tout le monde, mais je vais faire au mieux.

Je tiens d'abord à remercier Sandra Lavorel pour m'avoir encadré à sa façon, c'est à dire en 'totale liberté, mais surveillée'. Son dynamisme, son optimisme, sa lucidité et nos nombreuses discussions m'ont énormément apporté et m'ont fait mûrir et progresser plus vite que je n'aurais pu imaginer.

Merci aussi à Miguel Araújo et Lluís Brotons, qui auront été pour moi comme des seconds encadrants et surtout des amis.

Je tiens aussi à avoir une pensée pour toutes les personnes rencontrées durant mes collaborations et voyages, comme Guy Midgley, Martin Sykes, Wolfgang Cramer, Santi Sabate, Mathieu Rouget, Carles Gracia, Dinah Millar, Richard Pearson, Enrique Martinez, Paul Leadley et bien d'autres encore qui m'ont permis de m'ouvrir, d'apprendre à écouter et qui je pense ont contribué largement à ce travail.

Un remerciement particulier pour Joan Pino et Jordi Vayreda qui m'ont accueilli en Catalogne pour une collaboration plus que fructueuse et une amitié solide. De même que Guy Midgley, Dinah Millar et Mathieu Rouget avec qui ces deux mois passés en Afrique du sud auront été plus qu'enrichissants.

Enfin un grand merci à Sébastien Lavergne pour son interaction, sa clairvoyance et son énorme investissement à lire plusieurs versions préliminaires longues et fastidieuses de ce manuscrit.

Merci bien évidemment aux différentes personnes qui ont participé aux conseils de thèse comme Max Debussche, Jean-Louis Martin, John Thompson, Daniel Chessel, Joan Pino et Isabelle Chuine. Une pensée spéciale pour Max, Jean-Louis et Eric Garnier pour leur soutien durant ces trois années, et à Daniel Chessel qui m'a permis de me retrouver à Montpellier et dont ses conseils me seront toujours bénéfiques.

Merci encore aux membres du jury, Antoine Guisan, Jaume Terradas, Joel Guiot, Philippe Choler et Doyle McKey, pour leurs commentaires plus qu'utiles et leur disponibilité.

Pour finir un grand merci aux potes du restaurant et du café qui sans le savoir m'ont permis de m'évader chaque jour, grâce à ces discussions aussi diverses que sur la musique ou la politique. Donc merci à Sophie, POC, Seb, Bodil, Lluís, Adeline, Justin, Paul, François, Emilie, Clélia, George, Fabiana, Gabriel, Sandrine, Axel, Ophélie et j'en oublis.

Un remerciement spécial pour Karine, Maryse ou Anne-Marie de la direction du CEFÉ, toujours heureuse de signer une mission deux jours avant de partir, et surtout pour leur éternelle gentillesse, bonne humeur et efficacité. De même, merci au CEFÉ de m'avoir accueilli et à toutes les personnes que j'ai pu oublier.

Enfin un immense remerciement à Fanny pour avoir supporté mes week-ends passés au CEFÉ, pour ses relectures fastidieuses de ce manuscrit en feignant l'intéressement, et surtout pour cette leçon d'abnégation qu'elle me donne chaque jour.

Avant propos

Cette étude a bénéficié du soutien financier de la Commission Européenne sous la forme d'un contrat de travail de 3 ans à travers le projet européen ATEAM (Advanced Terrestrial Ecosystem Analysis and Modelling) dirigé par Wolfgang Cramer (Postdam Institute, Allemagne). L'encadrement scientifique et le support logistique ont été assurés par le Centre d'Ecologie Fonctionnelle et Evolutive de Montpellier (UMR CNRS).

Cette étude a aussi été aidée financièrement par le Laboratoire Européen Associé Montpellier-Barcelone «Dynamique des écosystèmes méditerranéens dans un monde changeant», permettant une collaboration enrichissante avec le Centre de Recherche en Ecologie et Applications Forestières (CREAF) et particulièrement Joan Pino, Santi Sabate et Jordi Vayredas. Cette thèse a pu également compter sur le soutien financier du Ministère des Affaires Etrangères et de l'Enseignement Supérieur et la Fondation Nationale pour la Recherche (NRF) d'Afrique du Sud qui a permis un séjour de deux mois pour une collaboration avec Guy Midgley (National Botanical Insitute, Cape Town).

Le texte qui suit est composé de deux parties. La première propose la synthèse d'une grosse partie des travaux réalisés lors de cette thèse. La seconde est composée de 15 articles rédigés en anglais, qui présentent de manière détaillée et exhaustive les différentes approches et analyses employées dans cette thèse. Bien que la synthèse ait été écrite de façon à limiter le recours aux annexes, il sera parfois nécessaire au lecteur de se reporter aux articles cités en début de chaque étude. 9 des 15 articles utilisés dans ce manuscrit sont publiés dans des revues internationales à comité de lecture, 3 sont en cours d'évaluation, et 3 autres sont en phase de préparation.

Les articles ont été regroupés sous la forme de quatre annexes, correspondant respectivement aux parties 1, 2, 3 et Conclusion.

INTRODUCTION

CHANGEMENTS GLOBAUX ET BIODIVERSITE

1	DEFINITION DES CHANGEMENTS GLOBAUX	- 8 -
1.1	LES ENJEUX	- 8 -
1.2	ECHELLE ET HETEROGENEITE GEOGRAPHIQUE DES CHANGEMENTS GLOBAUX.....	- 8 -
2	REPONSES ECOLOGIQUES AUX CHANGEMENTS GLOBAUX RECENTS.....	- 9 -
2.1	CHANGEMENT DANS LA PHENOLOGIE ET LA PHYSIOLOGIE DES ESPECES	- 10 -
2.2	CHANGEMENT DE DISTRIBUTION DES ESPECES	- 10 -
2.3	CHANGEMENTS DE COMPOSITION DES ECOSYSTEMES ET DES INTERACTIONS INTERSPECIFIQUES.....	- 11 -
3	COMMENT EVALUER LES EFFETS DES CHANGEMENTS GLOBAUX SUR LA BIODIVERSITE ?.....	- 12 -
3.1	OBSERVATION	- 12 -
3.2	MODELISATION EXPERIMENTALE	- 13 -
3.3	MODELISATION EMPIRIQUE	- 13 -
4	OBJECTIFS GENERAUX DE LA THESE.....	- 14 -

1ERE PARTIE

PREDICTIONS DES PATRONS ACTUELS DE DISTRIBUTIONS

1	INTRODUCTION	- 16 -
2	HYPOTHESES ET APPROCHES EN BIOGEOGRAPHIE PREDICTIVE.....	- 17 -
2.1	CONCEPT DE NICHE ECOLOGIQUE	- 17 -
2.2	EQUILIBRE <i>VERSUS</i> NON-EQUILIBRE	- 18 -
2.3	ESPECES <i>VERSUS</i> ASSEMBLAGES.....	- 19 -
3	MODELISATION STATIQUE BASEE SUR LES NICHES	- 20 -
3.1	THEORIE DES MODELES	- 20 -
3.2	APPROCHES EN BIOGEOGRAPHIE PREDICTIVE	- 21 -
3.3	CHOIX DES MODELES	- 26 -
3.4	CHOIX DES ESPECES : IMPACTS DE LA CHOROLOGIE	- 32 -
3.5	CHOIX DES VARIABLES EXPLICATIVES	- 35 -

2^{EME} PARTIE

**PROJECTION DES FUTURS HABITATS POTENTIELS DE LA BIODIVERSITE :
APPROCHES ET METHODOLOGIES**

1	INTRODUCTION	- 47 -
2	OPTIMISATION DES PREDICTIONS ET FIABILITE DES PROJECTIONS.....	- 49 -
3	APPROXIMATION DE LA NICHE REALISEE DES ESPECES : CONSEQUENCE POUR LES PROJECTIONS SOUS CHANGEMENTS GLOBAUX	- 53 -
3.1	EFFETS DE LA RESTRICTION DE L'ENVELOPPE DES CONDITIONS CLIMATIQUES SUR LA FIABILITE DES MODELES.....	- 55 -
3.2	EFFET DE L'APPROXIMATION INCOMPLETE DE LA NICHE BIOCLIMATIQUE DES ESPECES.....	- 56 -
4	FUTURS HABITATS POTENTIELS DES ESPECES : LA PRISE EN COMPTE DE L'UTILISATION DES TERRES A POSTERIORI	- 57 -
5	CONCLUSION.....	- 60 -

3^{EME} PARTIE

SENSIBILITES DES ESPECES ET SENSIBILITE GEOGRAPHIQUE AUX CHANGEMENTS CLIMATIQUES : PROJECTIONS ET INCERTITUDES

1	INTRODUCTION	- 64 -
2	SENSIBILITE DES ESPECES AUX CHANGEMENTS CLIMATIQUES : EFFETS DES PROPRIETES DES NICHES ET DE L'EXPOSITION GEOGRAPHIQUE	- 65 -
2.1	EXPOSITION DES PLANTES EUROPEENNES AUX CHANGEMENTS CLIMATIQUES	- 66 -
2.2	LIMITATIONS METHODOLOGIQUES.....	- 67 -
2.3	EST-CE QUE LES ESPECES MARGINALES SONT LES PLUS SENSIBLES AUX CHANGEMENTS CLIMATIQUES ?.....	- 68 -
2.4	LES ESPECES SPECIALISTES SONT-ELLES PLUS SENSIBLES AUX CHANGEMENTS CLIMATIQUES QUE LES ESPECES GENERALISTES ?	- 69 -
2.5	EST-CE QUE LES ESPECES AVEC DES AIRES DE REPARTITIONS LIMITEES SONT PLUS SENSIBLES AUX CHANGEMENTS CLIMATIQUES QUE LES ESPECES LARGEMENT REPANDUES ?	- 69 -
3	SENSIBILITE DES POOLS REGIONAUX D'ESPECES AUX CHANGEMENTS CLIMATIQUES : PROJECTIONS ET INCERTITUDES	- 71 -
3.1	INCERTITUDES DES PROJECTIONS DANS LE FUTUR	- 72 -
3.2	SENSIBILITE DES POOLS REGIONAUX D'ESPECES PAR PIXEL AUX CHANGEMENTS GLOBAUX	- 75 -
4	CONCLUSION.....	- 83 -

CONCLUSIONS ET PERSPECTIVES

1	CONSIDERATIONS ECOLOGIQUES ET BIOGEOGRAPHIQUES.....	- 88 -
2	CONSIDERATIONS METHODOLOGIQUES.....	- 90 -
3	VERS UNE APPROCHE BASEE SUR LES GROUPES DE REPONSE	- 91 -
4	VERS UNE APPROCHE MULTI-EHELLE SPATIALEMENT EXPLICITE	- 93 -

BIBLIOGRAPHIE.....	-95-
---------------------------	-------------

Annexe 1

Prédictions des patrons actuels de distributions

- 1.1 Generalized models vs classification tree analysis: predicting spatial distributions of plant species at different scales.
Thuiller W., Araújo M.B., Lavorel S. *Journal of Vegetation Science* (2003) **14**, 669-680.
- 1.2 Evaluating species distribution models: methodologies and conservation implications of prevalence and probability thresholds
Thuiller W., Howell C.A., Midgley G.F., Rouget M *Soumis*
- 1.3 Presence-absence vs presence-only modelling methods for predicting bird habitat suitability
Brotons L., Thuiller W., Araújo M.B., Hirzel A.H. *Ecography* (2004) sous presse
- 1.4 Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain)
Thuiller W., Vaydera J., Pino J., Sabaté S., Lavorel S., Gracia C. *Global Ecology & Biogeography* (2003) **12**, 313-325.
- 1.5 Do we need land-cover data to model species distributions in Europe?
Thuiller W., Araújo M.B., Lavorel S. *Journal of Biogeography* (2004), sous presse
- 1.6 Downscaling European species atlas distributions to a finer resolution: implications for conservation planning
Araújo M.B., Thuiller W., Williams P.H., Reginster I. *Global Ecology and Biogeography* (2004) sous presse

Annexe 2

Projeter les futures habitats potentiels de la biodiversité : approches et méthodologies

- 2.1 BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change
Thuiller W. *Global Change Biology* (2003) **9**, 1353-1362
- 2.2 Effects of restricting environmental range of data to project current and future species distributions
Thuiller W., Brotons L., Araújo M.B., Lavorel S. *Ecography* (2004), sous presse
- 2.3 Developing regional and species-level assessments of climate change impacts on biodiversity: A preliminary study in the Cape Floristic Region

Midgley G.F., Hannah L., Millar D., Thuiller W., Booth A. *Biological Conservation* (2003) **112**, 87-97

Annexe 3

Sensibilité des espèces et sensibilité géographique aux changements climatiques : projections et incertitudes

- 3.1 Niche properties and geographic extent as predictors of species sensitivity to climate change

Thuiller W., Lavorel S., Araújo M.B. *Soumis*

- 3.2 Uncertainties in species range shifts projections under global climate change scenarios

Thuiller W., Araújo M.B., Brotons L., Lavorel S., Midgley G.F., Williams P.H. *En préparation*

- 3.3 Patterns and uncertainties of species' range shifts under global change

Thuiller W. *Soumis*

- 3.4 Impact of global climate change on distributions of European taxa: projections under a set of IPCC storylines

Thuiller W., Araújo M.B., Lavorel S., Williams P.H. *En préparation*

Annexe 4

Conclusions et perspectives

- 4.1 Consequences of land use change on rare plant occurrence and extinction in a Mediterranean region: a 115 years study

Lavergne S., Thuiller W., Molina J., Debussche M. *En préparation*

- 4.2 Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* species in the Cape Floristic Region

Thuiller W., Lavorel S., Midgley G.F., Lavergne S., Rebelo A.G. *Ecology* (2004), sous presse

INTRODUCTION :

CHANGEMENTS GLOBAUX ET BIODIVERSITÉ

“It is not necessary in any empirical science to keep an elaborate logicomathematical system always apparent, more than it is necessary to keep a vacuum cleaner conspicuously in the middle of a room at all time.

When a lot of irrelevant litter has accumulated, the machine must be brought out, used, and then put away”

G.E. Hutchinson, 1957, Concluding Remarks, p. 415

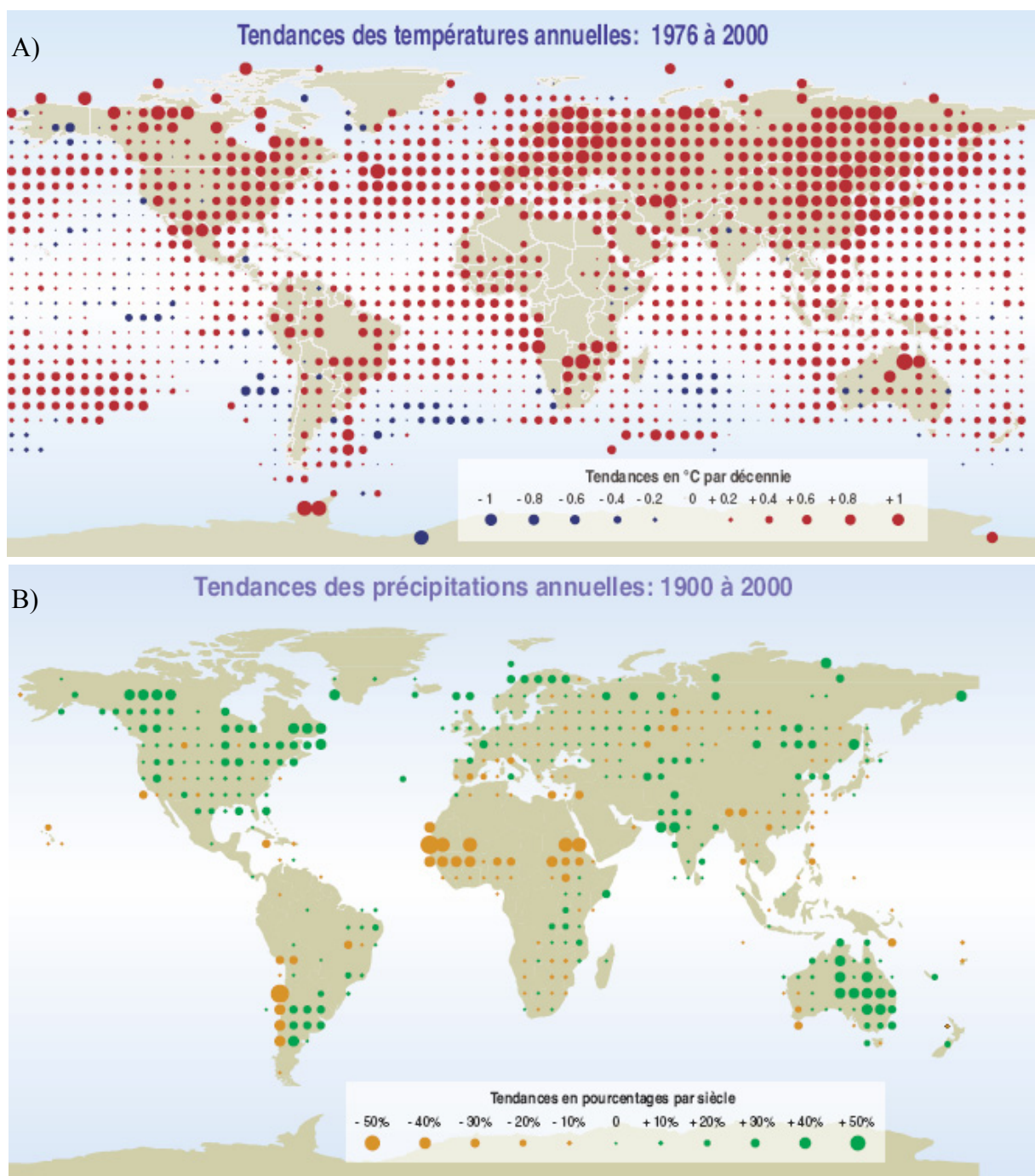
1 Définition des changements globaux

1.1 Les enjeux

Bien qu'il soit souvent ramené uniquement à des problèmes d'ordre climatique, le terme changement global peut se référer à une série de changements naturels ou d'origine anthropique sur la structure biologique et physique de la Terre, qui dans leur ensemble ont des effets significatifs à échelle globale (IPCC 1996). Depuis la conférence de Rio en 1992, il est admis que l'augmentation constante de la population mondiale (de 3,5 milliards dans les années soixante à 6 milliards à la fin du siècle) et de la consommation en énergie due au développement économique, ainsi que les changements consécutifs dans la composition de l'atmosphère, devrait changer globalement l'équilibre de la Terre. Les modèles numériques (GCM : modèles de circulation générale) calibrés sur les 100 dernières années, projettent que le réchauffement climatique devrait s'accroître dans les années à venir et que les températures pourraient augmenter de 1,4 C à 5,8 C dans les 100 prochaines années selon que l'atmosphère sera plus ou moins chargée en gaz à effet de serre (Figure 1). Cette fourchette assez large provient d'une part de l'incertitude des scénarios, et d'autre parts de la diversité des modèles pour un même scénario. Toutefois, si la température moyenne pourrait augmenter de 3°C, les réchauffements locaux se traduiraient par des augmentations allant jusqu'à 10°C dans certains cas (IPCC 1996). Bien que sur les temps géologiques de telles augmentations aient déjà été estimées, c'est la rapidité de ces changements qui inquiète. Comment les espèces et les écosystèmes vont-ils réagir à ces changements futurs alors que le réchauffement sur les 50 dernières années (1°C) semblent avoir déjà provoqué un nombre conséquent d'extinctions et de migrations d'espèces (Parmesan & Yohe 2003) ?

Divers programmes internationaux et nationaux ont donc été mis en place pour analyser et prédire l'impact des changements globaux sur les systèmes vivants. Dans ce cadre, deux grands types de changement global sont étudiés : les changements climatiques et les changements d'utilisation des terres. Ces deux notions sont très étroitement liées car les changements climatiques impliquent des modifications dans la gestion du sol, mais il est probable que l'augmentation de l'agriculture intensive et des pollutions associées, l'urbanisation croissante et le développement industriel influent fortement en retour le climat.

Figure 1 : A) Un réchauffement à grande échelle de la surface des terres et des océans a eu lieu au cours des vingt-cinq dernières années, les plus importantes élévations de température se situant à moyenne et haute latitude en Amérique du Nord, en Europe et en Asie. De grandes zones de refroidissement ont été observées uniquement dans certaines parties de l’océan Pacifique et de l’océan Austral et dans l’Antarctique. Le réchauffement plus rapide des terres par rapport aux océans correspond aux changements des variations climatiques naturelles observés, dans le cas, par exemple des Oscillations de l’Atlantique Nord et Arctique, et aux simulations de profils de réchauffement dû aux gaz à effet de serre. B) Pendant le XXe siècle, en moyenne, les précipitations ont augmenté sur les continents à l’extérieur des tropiques, mais ont diminué dans les zones désertiques d’Afrique et d’Amérique du Sud. Bien que les données climatiques indiquent une augmentation générale des précipitations, cohérentes avec l’élévation de la température et l’augmentation de l’humidité atmosphérique, les tendances en matière de précipitations sont extrêmement variables entre les régions et, pour certaines régions continentales, ne sont documentées que pour le XXe siècle. Au cours des cent dernières années, on observe des tendances à long terme relativement réduites dans les zones terrestres de sécheresse ou de forte humidité, mais dans de nombreuses régions ces changements sont dominés par une variabilité climatique interdécennale et multidécennale pour laquelle il n’existe pas de tendance évidente pendant le XXe siècle. Adapté de Watson et. al (2001)



1.2 Echelle et hétérogénéité géographique des changements globaux

La prise en compte de l'échelle et de la variabilité spatiale est cruciale dans l'évaluation des changements globaux et de leurs impacts sur les systèmes naturels. Il existe une stratification latitudinale naturelle des températures des zones les plus chaudes (équatoriales) aux zones les plus froides (pôles). Les modèles climatiques projettent des changements d'une magnitude plus importante vers les pôles, ce qui tendrait à diminuer le déséquilibre climatique existant entre l'équateur et les pôles. La variabilité spatiale des changements est encore facilement discernable à échelle continentale ou régionale (Figure 1). A ces échelles, on observe une hétérogénéité non seulement latitudinale et longitudinale mais aussi continentale vs océanique, et altitudinale (plaines vs montagnes). L'impact des changements climatiques sur les écosystèmes et plus particulièrement sur les espèces sera donc différent selon leur position géographique. Les changements climatiques sont fréquemment assimilés à une simple augmentation de température et une baisse des précipitations, mais ce sont plutôt les combinaisons complexes de variables climatiques, topographiques et géologiques, qui sont susceptibles d'être modifiées et donc d'influencer la distribution des organismes.

L'impact des modifications d'utilisation des terres et de couverture végétale jouent aussi un rôle important sur la distribution des espèces, en modifiant les mosaïques paysagères (par ex. par fragmentation des habitats) et le climat. Cette influence est très hétérogène, diverse et variable dépendant de la localisation géographique, des pratiques ancestrales, et du contexte politico-économique de la zone étudiée. Tous ces facteurs interagissent et influencent à des échelles diverses et de façon spatialement hétérogène la biodiversité et la distribution des espèces.

2 Réponses écologiques aux changements globaux récents

Il y a maintenant des preuves empiriques difficiles à réfuter que les récents changements climatiques (ayant eu lieu au cours des 100 dernières années) affectent des organismes avec des distributions géographiques diverses (Hughes 2000; McCarthy 2001; Walther *et al.* 2002; Parmesan & Yohe 2003; Root *et al.* 2003). Les impacts des changements globaux sur les espèces et indirectement sur les écosystèmes peuvent être résumés en trois grandes classes (Figure 2).

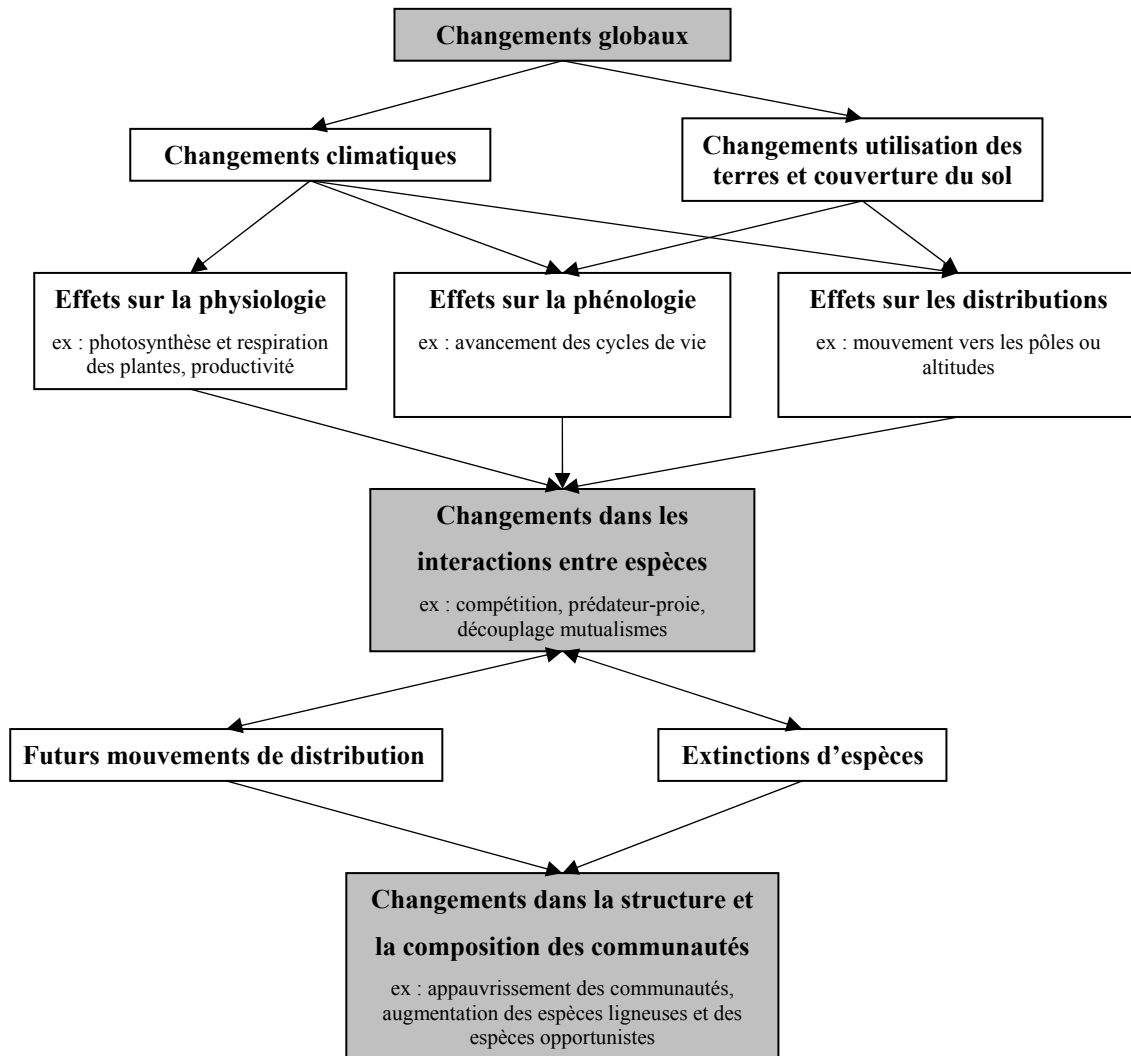
2.1 Changement dans la phénologie et la physiologie des espèces

La phénologie (date des activités saisonnières des animaux et des plantes) est peut-être le processus le plus simple pour appréhender les changements dans l'écologie des espèces en réponse aux changements globaux (Schwartz 2003). Différentes études, en Europe et en Amérique du Nord, ont montré des modifications phénologiques qui reflètent très probablement des réponses aux récents changements climatiques (Bairlein & Winkel 2001; Menzel & Estrella 2001). En général, les auteurs ont mis en évidence des changements dans les dates d'activités printanières, comme des dates de premières nidifications, de premiers chants d'oiseaux, d'arrivée des oiseaux migrateurs plus précoces, ou encore des dates de floraison précoces (Fitter & Fitter 2002). Certaines études indiquent aussi des événements phénologiques automnaux plus tardifs, mais ces modifications sont moins prononcées et montrent un patron plus hétérogène (Walther *et al.* 2002). En Europe, par exemple sur les cinquante dernières années, la coloration des feuilles montre un retard progressif de 0,3 à 1,6 jours par décennie, alors que la longueur de la saison de croissance a augmenté de 3,6 jours par décennie dans certaines régions. Même si les changements d'ordre phénologique entre les oiseaux et les plantes suivent les mêmes tendances (Bradley *et al.* 1999), ils ne sont pas toujours synchrones dans les différents groupes taxinomiques et peuvent avoir des conséquences écologiques variées (Visser *et al.* 1998; Visser & Holleman 2001). Des changements dans la physiologie des plantes, la productivité et la croissance ont aussi été observés (Hughes 2000). Par exemple, plusieurs études ont montré que la densité des stomates des plantes collectées récemment était significativement inférieure à celle des spécimens d'herbarium collectés 70 à 200 ans avant (Beerling & Kelly 1997), et que cette variation était fortement corrélée à l'augmentation des niveaux de CO₂ du siècle dernier. Des tendances positives et négatives dans l'abondance du phytoplancton indiquent que les changements climatiques pourraient aussi affecter la productivité dans les océans, bien que de grandes différences régionales soient apparentes (Reid *et al.* 1998).

2.2 Changement de distribution des espèces

Les variables climatiques influencent incontestablement la distribution géographique des espèces, souvent au travers de seuils de tolérance physiologique à la température, la précipitation ou l'humidité, spécifiques à l'espèce. Les espèces migratrices sont parmi les plus étudiées mais montrent souvent de larges fluctuations annuelles dans leurs sites de nidifications, rendant difficile la distinction des modifications de distributions à long terme (Crick & Sparks 1999). Par contraste, les changements de limites de distribution pour les

Figure 2 : Effets possibles des changements globaux sur les communautés. Les espèces peuvent répondre de trois façons provoquant un changement dans les interactions biotiques. Ces changements peuvent mener à l'extinction d'espèces ou à de possibles mouvements dans les aires de distributions, provoquant un changement dans les communautés. Adapté de Hughes (2000)



espèces plus sédentaires découlent de processus d'extinctions et de colonisations plus simples à appréhender (Walther *et al.* 2002). Il est maintenant clair que des mouvements vers les pôles et vers des altitudes plus élevés sont observés dans de nombreux groupes taxinomiques et régions au cours du 20^{ème} siècle (McCarthy 2001; Parmesan & Yohe 2003). Les facteurs affectant la distribution des espèces interagissent de manière complexe, et il n'est pas surprenant que de simples corrélations avec la température ne soient pas toujours observées. Les modifications de distribution sont aussi souvent épisodiques plutôt que graduelles ou continues (Walther *et al.* 2002). Par exemple, dans les régions sous influence d'El Niño Southern Oscillation (ENSO), des changements de distributions peuvent avoir lieu très rapidement durant les épisodes de fortes chaleurs, avec des changements inverses durant les périodes plus froides. De plus les extrêmes climatiques sont des déterminants importants des changements actuels de distributions (Easterling *et al.* 2002). De la même manière les changements d'utilisation des terres influencent les distributions d'espèces et provoquent des migrations et/ou des extinctions (Hobbs 2000). Par exemple, Robinson & Sutherland (2002) ont montré que l'intensification de l'agriculture en Angleterre après la seconde guerre mondiale avait profondément affecté la distribution des oiseaux spécialistes de terres cultivées et provoqué la disparition de nombreux taxons. De plus, l'augmentation des flux mondiaux et du commerce international a provoqué une brèche dans les frontières biogéographiques et facilité la dispersion d'espèces à l'intérieur de régions qu'elles n'auraient naturellement jamais atteintes. Cette dispersion d'espèces, ajoutée aux modifications de l'utilisation des terres et de la couverture du sol, offre des opportunités pour la colonisation d'espèces ou l'expansion de populations invasives. Certaines espèces invasives arrivent à coloniser ces habitats perturbés et modifient les écosystèmes naturels créant ainsi des écosystèmes dits émergents (« Un écosystème ayant une composition d'espèces jamais observée dans aucun biome»; Hobbs, R.J. non publié).

2.3 Changements de composition des écosystèmes et des interactions interspécifiques

Les changements de phénologie, de physiologie et de distributions des espèces altèrent inévitablement la compétition et les interactions entre espèces, avec des possibles répercussions sur l'abondance locale des espèces et leurs aires de distributions (Hughes 2000). Ce stress physiologique et les interactions entre espèces devraient provoquer des extinctions d'espèces et promouvoir l'installation d'espèces opportunistes et invasives (Figure 2). Les assemblages de pollen à différentes périodes avant et après les dernières glaciations montrent

que les communautés avaient des compositions différentes d'aujourd'hui (Anderson & Brubaker 1994; Clarkson *et al.* 1995), suggérant que les communautés dans le futur ne devraient pas forcément être la réplique des communautés actuelles, mais qu'au contraire des assemblages jamais observés pourraient être formés. Par ailleurs, il a été noté, grâce aux enregistrements passés de végétation et des modèles de changements climatiques, que la sensibilité variable des espèces à des changements de milieu pouvait provoquer des trajectoires différentes des communautés (Brubaker 1986; Rupp *et al.* 2000).

De la même manière que pour les communautés, les changements climatiques perturbent le fonctionnement des écosystèmes. Par exemple, la propagation rapide des arbres fixateur d'azote durant les périodes post-glaciaires ont provoqué l'augmentation rapide de la productivité des écosystèmes aquatiques (Hu *et al.* 2001), et l'invasion des fixateurs d'azotes à l'intérieur des systèmes sans fixateur ont provoqué des augmentations substantielles de la productivité et de l'inflammabilité (Vitousek & Walker 1989). L'invasion par des graminées inflammables a augmenté l'intensité et le retour des feux dans des prairies semi-arides (Whisenant 1991). De tels changements de milieu et des régimes de perturbations dus à des espèces invasives influencent en retour les patrons futurs d'assemblages dans les communautés (Lavorel & Garnier 2002).

3 Comment évaluer les effets des changements globaux sur la biodiversité ?

3.1 Observation

Des observations à grandes échelles et sur de longues périodes de temps pourraient être réalisées pour mesurer efficacement les effets des changements climatiques tant au niveau de la phénologie ou de la physiologie des espèces, qu'au niveau des changements de distributions. Cependant, l'échelle humaine est trop courte et le temps d'observation trop long pour permettre ce genre d'analyses très coûteuses en temps et en main d'œuvre. Ces études peuvent être réalisées uniquement sur des espèces à temps de génération assez courts et répondant rapidement à des variations de leur milieu (ex : insectes ou micro-organismes)(Davis *et al.* 1998).

Les observations paléoécologiques (ex : pollen) permettent de comprendre les réponses de certaines espèces ou formes de vie aux réchauffements climatiques post-glaciaires (Blondel & Aranson 1999). Bien que difficilement extrapolable aux changements climatiques actuelles compte tenu des résolutions temporelles des analyses (~1000 ans), ces études permettent de

mettre en lumière les réponses individuelles des espèces aux modifications du milieu (Prentice *et al.* 1991), ou bien encore certains processus de spéciation apparus grâce à des zones refuges (Taberlet & Cheddadi 2002).

3.2 Modélisation expérimentale

Pour évaluer l'impact des changements climatiques sur les espèces, une modélisation expérimentale de la relation des espèces avec des variables climatiques peut être réalisée. Les installations de type « écosystèmes artificiels » permettent de contrôler de nombreuses variables physiques et d'étudier les réponses des certaines espèces en terme de paramètres physiologiques ou phénologiques (Lawton 1994). Ces expériences permettent aussi d'étudier les interactions interspécifiques. Cependant elles ne peuvent être réalisées que sur un nombre limité d'espèces et en fonction de facteurs sélectionnés comme importants par les expérimentateurs. De plus, comme il s'agit d'expérimentation, la transposition des résultats aux conditions naturelles n'est pas toujours évidente. Enfin ce type d'expérience ne permet pas de mesurer des paramètres démographiques (ex : croissance de population, risque d'extinction), qui sont directement liés aux changements de distribution (colonisation versus régression). Toutefois, ces analyses sont cruciales car elles permettent de prouver expérimentalement la réponse différentielle des espèces aux changements de conditions climatiques et surtout de pouvoir donner des mesures létales de certains facteurs limitants comme les températures extrêmes ou l'aridité.

3.3 Modélisation empirique

Bien que la modélisation expérimentale permette de déterminer précisément les réponses physiologiques et phénologiques d'espèces sélectionnées aux changements climatiques, il apparaît que la modélisation empirique est l'unique voie pour évaluer l'impact des changements globaux sur la biodiversité sur de vastes échelles spatiales, temporelles et taxinomiques. La modélisation empirique permet de simuler les réponses des espèces aux facteurs du milieu, de projeter les futures distributions d'espèces ou de biomes, ou encore de simuler la dynamique d'espèces dans un paysage fragmenté. De nombreux modèles sont à disposition, à choisir selon les objectifs fondamentaux de l'étude, les organismes à modéliser ou encore l'échelle et la résolution des données. Il est évident que le modèle utilisé pour simuler la dynamique d'une plante dans un paysage fragmenté, sera différent d'un modèle utilisé pour prédire la distribution de 2000 espèces à travers le monde et sur les 50 prochaines années.

Certains modèles utilisent les relations mesurées des espèces ou biomes à des variables de milieu en termes de physiologie ou de phénologie et tentent de produire des cartes de distributions où l'organisme modélisée peut ou ne peut pas exister actuellement et dans le futur, compte tenu de scénarios de changements climatiques (modèles mécanistes) (ex : Chuine *et al.* 2000; Kleidon & Mooney 2000; Ni *et al.* 2000). Certains modèles basés sur les mêmes données que précédemment, mais prenant en compte les capacités de dispersion et de colonisation des espèces ainsi que certains processus de compétition, permettent de simuler d'une manière dynamique les réponses des biomes à des modifications de leur milieu (modèles dynamiques)(ex : Bachelet *et al.* 2003; Sitch *et al.* 2003). Enfin, d'autres modèles tentent de reproduire les distributions actuelles des organismes étudiés à l'aide de variables de milieu observées permettant de déterminer des enveloppes dans lesquelles les espèces peuvent vivre, et projettent ces enveloppes dans le futur en utilisant des scénarios de changements sur les variables de milieu utilisées (modèles statiques)(ex : Gottfried *et al.* 1995; Guisan *et al.* 1998; Peterson *et al.* 2002).

Pour étudier l'impact des changements globaux à diverses échelles temporelles et spatiales, et sur un grand nombre d'espèces, la modélisation statique basée sur les niches, qui fait l'objet de cette thèse, semble être tout particulièrement appropriée.

4 Objectifs généraux de la thèse

L'objectif principal de cette thèse est d'évaluer, comprendre et prédire les impacts des changements globaux sur la distribution de la biodiversité à échelle européenne. Mon travail a donc été articulé autour de trois grands champs disciplinaires qui sont la modélisation biostatistique, l'écologie et la biogéographie. Ceci a impliqué une forte composante méthodologique permettant la construction et la compréhension des modèles, une composante écologique sur le choix des espèces et le choix des variables, ainsi qu'une composante plus biogéographique sur la compréhension des relations espèce-milieu, la prédiction de patrons à vastes échelles spatiales et l'interprétation biogéographique des résultats obtenus.

Le manuscrit s'articulera en trois grandes parties.

- La première partie traitera tout d'abord des concepts et approches liés à la modélisation des distributions actuelles d'organismes. Ensuite deux analyses comparatives sont présentées permettant de tester la fiabilité de certains modèles en fonction de l'échelle et de la résolution, ainsi que de la qualité des données initiales. Je présenterai aussi une analyse sur l'utilité des modèles pour prédire les habitats potentiels des espèces, mais aussi

leur difficulté sur certains types d'espèces avec des aires de distributions incomplètes. Pour finir une analyse de la pertinence des variables de milieu en fonction, de l'échelle et des espèces, sera effectuée. Ces quatre études permettent ainsi d'analyser les différents paramètres influençant la qualité et la fiabilité des prédictions des distributions d'espèces.

- La deuxième partie sera centrée sur l'application des modèles statiques dans un contexte de changement globaux. Tout d'abord, je présenterai succinctement ce que représentent et sur quoi sont basés les scénarios de changements climatiques. Ensuite une étude sera présentée concernant un outil que j'ai développé pour optimiser la prédiction des distributions actuelles d'organismes et permettre la projection des distributions dans le futur. Comme la projection des distributions des espèces dans le futur ne se fait pas sans risque, je présenterai une étude concernant les biais prédictifs des modèles liés à la restriction des limites bioclimatiques des organismes et aux organismes en limite de distributions dans la zone d'étude dans un contexte de changements globaux. Enfin, une dernière analyse exposera une approche séquentielle pour prendre en compte les modifications de l'utilisation des terres dans le futur pour prédire les futurs habitats des espèces.
- La troisième partie permettra, à l'aide des développements méthodologiques apportés par les deux premières parties, tout d'abord d'évaluer la sensibilité des espèces européennes aux changements climatiques, ceci en fonction de leurs propriétés écologiques. Outre la sensibilité des espèces, je présenterais aussi deux études sur la sensibilité du pool régional d'espèces présent par pixel en fonction d'une série de scénarios de changements climatiques. La première étude analysera l'incertitude des projections dans le futur, fonction d'une part, de paramètres méthodologiques (modèles), et d'autre part, de la variabilité dans les scénarios de changements climatiques. La deuxième étude permettra d'évaluer la sensibilité du pool régional d'espèces par pixel au niveau européen, par pays ainsi que par région biogéographique, et d'autre part de délimiter les zones géographiques où le turnover d'espèces pourraient être le plus important et provoquer des migrations.

1^{ERE} PARTIE

PREDICTIONS DES PATRONS ACTUELS DE DISTRIBUTIONS

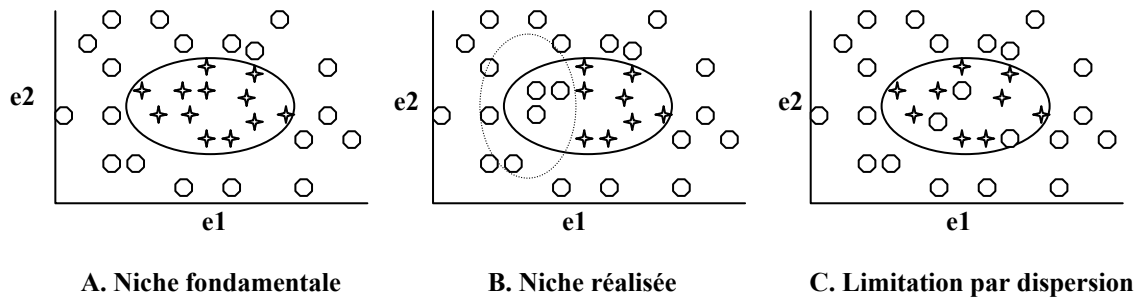
1 Introduction

L'importance du climat pour expliquer la distribution des plantes et animaux a été reconnue et étudiée au début du 19^{ème} siècle, notamment avec les travaux de van Humboldt & Bonpland (1807) et de Candolle (1855). La quantification de ces relations espèce-milieu représente le fondement de la biogéographie prédictive en écologie. Les modèles cherchent généralement à répondre à de nombreuses hypothèses ou questions, telles que "comment les facteurs de milieu contrôlent la distribution des espèces et des communautés ?". Outre un outil de recherche en autoécologie, la biogéographie prédictive est utilisée pour évaluer l'impact du changement d'utilisation des terres sur la distribution des organismes (Robinson *et al.* 1992; Kirilenko *et al.* 2000; Robinson *et al.* 2001), pour tester des hypothèses biogéographiques (Austin & Smith 1989; Mourelle & Ezcurra 1996; Leathwick & Austin 2001; Leathwick 2002), pour améliorer la qualité des atlas floristiques et faunistiques (Barbosa *et al.* 2003), pour mettre en place des réseaux de conservation (Araújo *et al.* 2002a; Araújo *et al.* 2002b) ou encore pour évaluer l'impact des changements climatiques sur les organismes (Bakkenes *et al.* 2002; Peterson *et al.* 2002; Sitch *et al.* 2003).

Il existe une grande variété de modèles utilisés pour décrire, expliquer, et prédire la distribution spatiale des espèces, allant des plantes terrestres (Austin *et al.* 1990; Saetersdal 1994), aux animaux aquatiques (Lek *et al.* 1996). De même différents modèles ont été implémentés pour étudier les communautés de plantes (Brzeziecki *et al.* 1995), les types de végétation (Brown 1994), les types fonctionnels, biomes ou assemblages complexes (Prentice *et al.* 1992; Lavorel *et al.* 1999). Ces modèles reposent sur des hypothèses et algorithmes différents rendant leur utilisation fortement dépendante des objectifs de l'analyse et de la disponibilité des données.

Cette première partie sera d'abord axée sur les approches conceptuelles et les hypothèses liées à l'utilisation de modèles pour comprendre et prédire la distribution actuelle des organismes. Plus particulièrement les hypothèses liées aux niches écologiques des espèces et à l'équilibre des espèces avec leur milieu seront détaillées. Comme explicité dans l'introduction et suivant les contraintes en données et les objectifs, j'ai choisi dans cette thèse d'utiliser un certain type de modèles dits basés sur les niches. La suite de cette partie sera donc consacrée à l'étude de ces modèles. Premièrement, je présenterais deux études comparatives, l'une testant la puissance prédictive de trois modèles couramment utilisés en biogéographie prédictive en fonction de l'échelle et la résolution des données, et l'autre testant la puissance prédictive de

Figure I.1 Niche écologique et distribution d'un organisme. Dans chaque diagramme, l'ellipse représente la niche fondamentale, englobant la combinaison des facteurs environnementaux (e1 et e2) pour lesquels l'espèce a un taux de croissance égal ou supérieur à 1. Les croix montrent la présence de l'espèce dans un habitat caractérisé par des valeurs particulières de e1 et e2, et les bulles indiquent l'absence de l'espèce dans un habitat. A) Espèce présente là où les conditions sont favorables ; B) Espèce absente dans les portions de sa niche fondamentale occupées par un compétiteur dominant ; C) Espèce absente de certains habitats favorables à cause d'un faible pouvoir de dispersion, d'extinction locale ou du temps nécessaire pour leur colonisation. (Adapté de Pulliam 2000)



deux modèles en fonction de la qualité des données. Ces deux premières études permettent d'analyser les avantages et inconvénients méthodologiques des modèles. Le reste de cette partie sera basée sur les paramètres écologiques qui peuvent influencer la fiabilité des modèles. D'une part, je présenterais une étude cherchant quels peuvent être les déterminants climatiques de la distribution de certains arbres à échelle régionale et les problèmes d'ajustement des modèles liés à l'appartenance phytogéographique des espèces étudiées. D'autre part, j'étudierai l'importance des variables de milieu sur la fiabilité des modèles. Plus particulièrement, j'étudierais l'influence de variables comme l'utilisation des terres et la couverture du sol comparée à des variables bioclimatiques pour prédire la distribution d'un grand nombre d'espèces à échelle européenne.

2 Hypothèses et approches en biogéographie prédictive

2.1 Concept de niche écologique

Expliquer et/ou prédire la distribution d'un organisme en fonction de paramètres du milieu repose sur le concept de niche écologique. Joseph Grinnell et Charles Elton ont introduit indépendamment au début du 20^{ème} siècle le concept de niche pour désigner une place ou un rôle dans un environnement ou une communauté (Grinnell 1917; Elton 1927). La différence majeure entre ces deux précurseurs était que le concept de Grinnell reposait sur l'environnement, et celui d'Elton sur la communauté. Dans ces fameuses « Concluding Remarks », Hutchinson (1957) donnait une nouvelle orientation et formulation au concept de niche qui est devenu depuis le point central de beaucoup de raisonnements et théories écologiques (Pulliam 2000). Hutchinson définit la niche fondamentale d'un organisme comme un hypervolume à n -dimensions, dans lequel à chaque point correspond un état du milieu permettant à l'organisme d'exister indéfiniment (Leibold 1995). Ce concept d'hypervolume définit une approche simple et rigoureuse pour quantifier la niche d'un organisme. Cette interprétation est décrite en Figure I.1-A : une espèce se situe dans n'importe quel site tant que les conditions y sont favorables (croix) et jamais où elles ne le sont pas (cercle ouvert). Cette représentation indique que sous conditions normales de reproduction et de dispersion, l'espèce est supposée occuper une région géographique qui est directement congruente avec la distribution de sa niche écologique. Etant supposé que dans la nature, des paramètres biotiques (ex : compétition ou interaction) et physiologiques (ex : reproduction ou dispersion) peuvent encore interférer, Hutchinson définit aussi la niche réalisée (Hutchinson 1957). La niche réalisée d'une espèce, celle réellement occupée par l'organisme, est plus petite que la niche

fondamentale car elle ne comprend que les portions de la niche fondamentale que l'organisme occupe réellement. Selon Hutchinson, il résulte de l'exclusion compétitive que l'organisme peut être absent de portions de sa niche fondamentale (Figure I.1-B). On peut ajouter à ce concept de niche réalisée que de nombreux organismes ont une capacité de dispersion limitée (Cain *et al.* 1998; Clark 1998), faisant qu'ils n'atteignent pas toujours des habitats potentiellement adéquats (Figure I.1-C).

Cette distinction entre niche fondamentale et niche réalisée est essentielle car elle préfigure de l'utilisation d'un type de modèle particulier. On distinguera les modèles mécanistes qui tentent de donner une représentation de la niche fondamentale (décrivant explicitement des processus démographiques et écologiques des organismes), des modèles basés sur les niches ('*Niche-based models*') qui simulent directement la niche réalisée des organismes, s'appuyant sur les relations observées espèce-milieu.

2.2 Equilibre *versus* non-équilibre

Parmi les approches disponibles pour expliquer et prédire la distribution des organismes, les modèles, dits statiques ou à l'équilibre, reposent sur le postulat que les espèces sont à l'équilibre (ou quasi-équilibre) avec leur milieu au moment où elles sont modélisées (voir les revues de Franklin 1995; Guisan & Zimmermann 2000). Ne pouvant distinguer la réponse transitoire de celle à l'équilibre d'un organisme en réponse à une variation stochastique et dynamique du milieu, ces modèles statiques sont conceptuellement incapables de traiter des situations de non équilibre (Prentice *et al.* 1992). L'état d'équilibre est souvent considéré comme une hypothèse nécessaire pour la prédiction à grande échelle. A grande échelle (Europe, Amérique du Nord), certains auteurs ont montré que l'hypothèse d'équilibre permettait d'expliquer les mouvements de certains grands biomes depuis les dernières glaciations (Huntley 1990; Prentice *et al.* 1991). Ces auteurs suggèrent que ces biomes à l'équilibre avec leur milieu ont répondu au réchauffement post-glaciaire de l'Holocène par des migrations à grandes échelles permettant d'expliquer leurs distributions actuelles. Toutefois, ils suggèrent aussi qu'il peut y avoir un temps de réponse pouvant aller jusqu'à 1000 ans due à la longévité des espèces ainsi qu'à leur plasticité phénotypique. A échelle Européenne et dans cette thèse, on supposera donc que la majorité des espèces sont en quasi-équilibre avec leur milieu, au moins dans leur limite d'aire de distribution. Cependant, il est évident que dans des situations plus locales à l'intérieur de leur limite d'aire de distribution, les espèces peuvent être en situation de non équilibre. Par définition, les situations avec de fortes perturbations, de fortes influences humaines, ou une dynamique rapide de succession seront modélisées avec

difficultés (Zimmermann & Kienast 1999) car il est très vraisemblable que la plupart des organismes sont dans des phases de réponse transitoires aux variations de milieu et non à l'état d'équilibre. Les espèces invasives actuellement ne sont donc elles aussi pas conceptuellement à l'équilibre puisqu'en train de coloniser des habitats favorables. Parmi les modèles à l'équilibre, on retrouvera certains modèles mécanistes (Kleidon & Mooney 2000; Ni *et al.* 2000) et tous les modèles corrélatifs, statistiques ou basés sur les niches (Iverson & Prasad 1998; Leathwick *et al.* 1998; Duckworth *et al.* 2000b) qui ne prennent donc pas en compte les phénomènes de compétition, succession, perturbation et les distributions transitoires. L'alternative à de tels modèles sont les modèles dynamiques (Korzukhin *et al.* 1996), qui modélisent les distributions d'organismes sur un pas de temps donné, en tenant compte par exemple des processus de compétition, de la capacité de dispersion et des taux de croissances des espèces, ou encore de la productivité du milieu (Sitch *et al.* 2003). Contrairement aux modèles à l'équilibre, les modèles dynamiques réclament de grosses quantités de données aussi bien sur l'écologie, la physiologie et la phénologie des organismes étudiés, que sur la structure du paysage, du sol et de l'évolution du climat à intervalles de temps réduits (Bachelet *et al.* 2003). Ce besoin limite actuellement ces modèles pour la prédiction d'un grand nombre d'espèces où la connaissance sur les paramètres écologiques, physiologiques et phénologiques est restreintes. Pour palier ce problème, les modèles dynamiques sont fréquemment utilisés sur les plantes et plus particulièrement sur la notion de types fonctionnels censés prendre en compte, d'une manière générale, la variété de structures et de fonctions des plantes (Smith *et al.* 1997). On peut citer par exemple, les herbacées tropicales et tempérées, ou bien les forêts de caducifoliés (Smith *et al.* 2001). L'avantage majeur de l'utilisation des types fonctionnels est qu'ils permettent de réduire la quantité de paramètres et de données nécessaires en étudiant la relation générale d'un type fonctionnel (et non pas de nombreuses espèces) avec le climat (Prentice *et al.* 1992; Ni *et al.* 2000). Le principal inconvénient est qu'il n'est pas certain que les espèces individuelles au sein d'un type fonctionnel répondent de la même manière aux changements de milieu.

2.3 Espèces *versus* assemblages

Une discussion importante dans l'analyse des patrons de distribution de la biodiversité oppose la vue "Gleasonienne", basée sur l'espèce, et la vue "Clemenstienne", basée sur les assemblages d'espèces (Prentice *et al.* 1992). Les données paléoécologiques attestent que les assemblages n'ont jamais été stables dans le passé (Webb 1981; Prentice 1986) et que les espèces répondent de manière individuelle aux changements de milieu (). Ceci justifie

l'analyse individuelle des espèces plutôt que celle des assemblages. Les assemblages modernes ont une histoire courte et il est peu probable qu'ils évoluent comme une seule entité sous l'effet des changements globaux (Huntley & Webb 1988; Birks 1993). Il a en effet été montré que des taxons qui ont aujourd'hui des aires de répartition chevauchantes ou presque identiques avaient des répartitions totalement disjointes durant l'Holocène (Huntley & Prentice 1988). Ainsi la composition des assemblages évoluerait au gré des changements de milieu. De ce fait l'utilisation de la modélisation statique et plus particulièrement la modélisation basée sur les corrélations entre distributions des organismes et variables de milieu ne devraient pas utiliser les assemblages d'espèces pour prédire leurs futures distributions.

Il existe donc différentes approches en modélisation pour comprendre et prédire l'impact des changements globaux sur la diversité. Ces approches sont basées principalement sur la distinction entre niches fondamentales *versus* réalisées, et équilibre *versus* non équilibre. Pour évaluer l'impact des changements globaux sur la biodiversité Européenne, c'est à dire sur un grand nombre d'espèces (plantes, mammifères, oiseaux, reptiles et amphibiens), j'ai choisis, compte tenu des disponibilités des données, une modélisation à l'équilibre basée sur les niches réalisées.

3 Modélisation statique basée sur les niches

3.1 Théorie des modèles

Le but de la biogéographie prédictive n'est pas seulement de prédire, mais aussi de comprendre les patrons observés. De façon un peu schématique, la pratique de la modélisation est associée à trois objectifs majeurs non exclusifs.

- Descriptif : Cet objectif vise à chercher de manière exploratoire les liaisons entre la variable réponse (une entité) et d'autres variables potentiellement explicatives, qui peuvent être nombreuses afin, par exemple, d'en sélectionner un sous-ensemble. En biogéographie prédictive, cet objectif est peu fréquent, car ce champ disciplinaire utilise un savoir acquis par d'autres champs disciplinaires (ex : écologie fonctionnelle ou des communautés).
- Explicatif : Cet objectif est sous-tendu par une connaissance a priori du domaine concerné et dont les résultats théoriques peuvent vouloir être confirmés, infirmés ou précisés par l'estimation de paramètres. Ceci revient donc à tester les effets d'une ou

Tableau I.1 : Clef pour la sélection d'un type de modèle, basée sur ses particularités et sur l'objectif de l'analyse.

1. Explicatif		
1.1. Multivarié		
1.1.1. Pondération des sites riches en espèces et relation unimodale entre variable réponse et gradient		CCA
1.1.2. Pondération uniforme des sites		
1.1.2.1. Relation linéaire entre variable de réponse et gradient		RDA
1.1.2.2. Pas d'hypothèse sur les relations		OMI
1.2. Univarié		
1.2.1. Variables réponse en présence/absence ou abondance		
1.2.1.1. Méthodes paramétriques		GLM, PLS
1.2.1.2. Méthodes non paramétriques		
1.2.1.2.1. Relations additives entre var. explicatives		GAM, Loess
1.2.1.2.2. Relations de tous types + interactions		CART
1.2.2. Variable réponse en présence seulement ou abondance		ENFA
2. Prédicatif		
2.1. Variables explicatives géographiques, variable réponse en présence seulement		Convex hulls, Kernel
2.2. Variables explicatives de milieu		
2.2.1. Variables réponse en présence/absence ou abondance		
2.2.1.1. Multivarié		CCA, RDA
2.2.1.2. Univarié		
2.2.1.2.1. Paramétrique		GLM
2.2.1.2.2. Non paramétrique		
2.2.1.2.2.1. Facilité d'interprétation		GAM, CART
2.2.1.2.2.2. Boîte Noire		ANN, GA
2.2.2. Variable réponse en présences seulement		
2.2.2.1. Poids similaires à toutes les variables		BIOCLIM, DOMAIN
2.2.2.2. Poids des var. expl. fonction de l'importance sur la var. rep.		ENFA

CCA (Canonical Correspondence Analysis : ter Braak 1986), RDA (Redundancy Analysis : Rao 1964), OMI (Outlying Mean Index : Doledec *et al.* 2000), GLM (Generalised Linear Models : McCullagh & Nelder 1989), PLS (Partial Least Square regression), GAM (Generalised Additive Models : Hastie & Tibshirani 1990), Loess (Locally Weighted Regression : Cleveland & Devlin 1988), CART (Classification And Regression Tree analysis : Breiman *et al.* 1984), ENFA (Ecological Niche Factor Analysis : Hirzel *et al.* 2002), ANN (Artificial Neural Networks : Openshaw & Openshaw 1997), GA (Genetic Algorithm : Openshaw & Openshaw 1997).

plusieurs variables de milieu sur une variable de réponse et pouvoir générer des tests d'hypothèses.

- Prédicatif : Ici l'accent est mis sur la qualité des estimateurs et des prédicteurs qui doivent, par exemple, minimiser une erreur quadratique moyenne ou maximiser une statistique d'adéquation spatiale. Ceci revient donc à rechercher des modèles parcimonieux, c'est à dire avec un nombre volontairement restreint de variables explicatives. Le meilleur modèle obtenu peut donner des estimateurs légèrement biaisés au profit d'un compromis pour une variance plus faible. Un bon modèle, comme on le verra par la suite, n'est donc pas forcément celui qui explique le mieux les données au sens d'une déviance minimale ou d'un R^2 au prix d'un nombre important de variables pouvant introduire des colinéarités. Le bon modèle est celui qui conduit aux prédictions les plus fiables.

En biogéographie prédictive, il est essentiel de comprendre et d'expliquer pourquoi une variable est un bon prédicteur. On cherchera donc un compromis entre la modélisation explicative et prédictive pour que les prédictions ne dépendent pas uniquement de propriétés mathématiques.

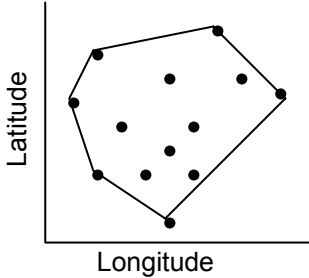
3.2 Approches en biogéographie prédictive

Les principales approches en modélisation statique peuvent être regroupées en sept grandes classes non exhaustives :

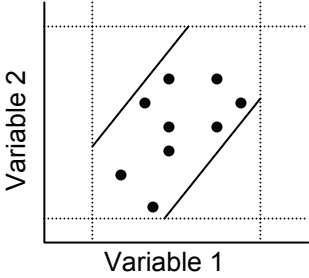
- Enveloppes géographiques
- Enveloppes climatiques
- Méthodes multivariées
- Outils de régressions
- Méthodes basées sur les arbres de décisions
- Méthodes d'apprentissages
- Autres méthodes

Ces méthodes diffèrent par le type de données nécessaires, leurs hypothèses sous-jacentes et les biais probables des résultats. Par conséquence, différentes méthodes peuvent prédire des habitats favorables contrastés pour une même espèce, bien que cela soit théoriquement un biais de prédiction. Les sections suivantes présentent ces 6 grandes classes et les propriétés qui les caractérisent. Dans le Tableau I.1 je propose une méthode qui permet par des critères simples de classer les modèles en fonction du but de l'étude (prédominance explicative ou prédictive), de leurs qualités et biais respectifs et de la structure des données écologiques (présence-absence vs présence seulement).

Figure 1.2 : Enveloppe géographique de type 'convex hulls' et enveloppe climatique.



A) Convex hulls



B) Enveloppe climatique

3.2.1 Enveloppes géographiques

Les enveloppes géographiques corrélient la distribution observée des espèces à des coordonnées géographiques. Par exemple les ‘convex hulls’ définissent l’étendue des points d’occurrence des espèces comme une zone contrainte qui peut englober tous les sites connus où l’espèce est présente (Figure I.2-A). L’inconvénient est que la contrainte de convexité sur les surfaces extrêmes induit une enveloppe dont le niveau de résolution est très grossier, provoquant une surestimation de la répartition des espèces (Ostro *et al.* 1999). Ces méthodes sont très sensibles à la qualité des données, aux données manquantes, et aux erreurs spatiales. Elles sont cependant utiles pour estimer l’aire de répartition de tout organisme recensé spatialement, mais pas pour des cartes plus précises, car elles englobent généralement une quantité de sites qui ne sont pas favorables pour l’organisme étudié.

3.2.2 Enveloppes climatiques

Tout comme les enveloppes géographiques, les enveloppes climatiques utilisent uniquement la présence de l’organisme considéré et ont donc l’avantage de pouvoir utiliser des données telles que des enregistrements ponctuels, des collections de musées ou toutes autres formes de relevés géoréférencés. Ces méthodes, comparées aux précédentes, utilisent le climat plutôt que la localisation géographique comme variable explicative, traduisant ainsi les processus écologiques sous-jacents qui affectent la distribution des organismes (Figure I.2-B). Par exemple, dans BIOCLIM, les présences observées d’une espèce sont utilisées en conjonction avec des données climatiques pour construire le profil climatique (enveloppe climatique) de l’espèce (Busby 1991). Une carte d’habitat favorable pour l’espèce sélectionnée peut ainsi être construite en classant chaque site selon sa position dans le profil de l’espèce. Une contrainte majeure est que donner un poids similaire à toutes les variables climatiques implique des hypothèses *a priori* improbables que toutes les variables peuvent avoir le même effet sur la distribution des espèces. Un autre exemple est DOMAIN (Carpenter *et al.* 1993) qui applique une mesure de distance multivariée pour créer des cartes d’habitats. DOMAIN définit une similarité environnementale calculée suivant la métrique de Gower (Legendre & Legendre 1998) entre un site auquel on s’intéresse et le site le plus similaire où l’espèce étudiée a été observée.

3.2.3 Méthodes multivariées

Toutes les méthodes de cette classe provenant de l'écologie des communautés sont peu flexibles. Par exemple, des méthodes qui supposent une relation fixe entre les espèces et leur milieu (relation gaussienne pour l'analyse canonique des correspondances), ou qui ne permettent pas d'inclure des interactions entre facteurs sont probablement peu utilisables pour prédire la distribution d'espèces à faible résolution spatiale (Elith & Burgman in press). Une méthode très connue mais peu utilisée en biogéographie prédictive est l'Analyse Canonique des Correspondances (CCA, ter Braak 1986). La CCA suppose une relation unimodale des espèces avec les variables de milieu utilisées et une amplitude écologique égale des espèces (Austin 2002). Cette hypothèse est peu réaliste et il n'est pas clair que la méthode soit robuste à sa violation (Jongman *et al.* 1995). De plus la CCA étant à l'origine construite pour l'écologie des communautés, elle est basée sur des combinaisons de facteurs de milieu (ordination) qui sont les mêmes pour toutes les espèces (Guisan *et al.* 1999). Dans la même catégorie que la CCA, on peut citer l'analyse de redondance (RDA : Redundancy Analysis). Cette méthode, qui est l'extrapolation canonique de l'Analyse en Composantes Principales (PCA), est similaire à une régression linéaire multivariée avec seulement des termes linéaires. Ceci implique une relation obligatoirement linéaire entre la distribution des espèces et les gradients sélectionnés ce qui limite son utilisation à des gradients courts ou tronqués, ou à des facteurs fortement limitants (ter Braak & Prentice 1987; Doledec *et al.* 2000).

3.2.4 Outils de régression

Parmi les méthodes de régression les plus utilisées en biogéographie prédictive, on peut citer les modèles linéaires généralisés (GLM) et les modèles additifs généralisés (GAM). Les GLM sont une grande famille incluant les cas particuliers des régressions linéaires ou de l'analyse de variance. Tous les GLM sont basés sur une variable dite de réponse (ex : présence ou abondance d'espèce, richesse spécifique ou présence d'un type fonctionnel), des variables prédictives (variables de milieu) et une fonction lien qui décrit la relation entre la valeur attendue de la réponse et les variables prédictives. Une grande variété de modèles peuvent être construits, dépendant des variables réponses (Binomiale ou Poisson) et fonctions liens (logit, log-vraisemblance) (Guisan 2002). Les GLM permettent de modéliser des relations paramétriques complexes entre les variables réponses et prédictives (ex : linéaires, quadratiques ou polynomiales), et de tester les interactions entre variables prédictives (Guisan *et al.* 1999).

Les GAM peuvent être considérés comme la forme non paramétrique des GLM dans lesquels la fonction spécifiant la relation entre la variable réponse et le(s) prédicteur(s) est remplacée par une fonction de lissage dépendante des données. Parmi la variété de fonctions de lissage disponibles (Venables & Ripley 2002), les ‘smoothing splines’ (cubic β -splines) sont les plus communément utilisées en biogéographie (Austin & Gaywood 1994; Shao & Halpin 1995; Lehmann *et al.* 2002). Les GAM sont très attrayants parce que la réponse n’est pas limitée par une fonction paramétrique, ce qui implique que la surface de réponse modélisée est une représentation plus réaliste de la réalité (Yee & Mitchell 1991).

Selon les données initiales, GLM et GAM estiment les probabilités de présence ou l’abondance d’un organisme. Les courbes de réponse modélisées entre les variables utilisées et la probabilité de présence ou abondance de l’espèce permettent d’explicitier les relations espèce-milieu. Ces deux types de modèles peuvent être utilisés pour projeter la distribution d’espèces modélisées dans d’autres sites non inclus dans l’analyse d’origine ou selon des changements de milieu par modifications des valeurs des variables de milieu dans les modèles (Beerling *et al.* 1995; Guisan *et al.* 1998).

3.2.5 Méthodes basées sur les arbres de décision

Les méthodes reposant sur les arbres de classification ou de régression (CART : classification tree and regression tree analysis) ont été introduites par Breiman *et al.* (1984) et Clark & Pregibon (1992). Un arbre est construit par scissions répétées des données, suivant une simple règle basée sur les variables de milieu. A chaque scission, les données sont partitionnées en deux groupes exclusifs aussi homogènes que possible. L’hétérogénéité d’un nœud est défini par la notion de déviance qui peut être interprétée comme une déviance d’un modèle Gaussien (arbre de régression) ou multinomial (arbre de classification) (Rouget *et al.* 2001). Pour contrôler la longueur de l’arbre, des fonctions d’élagage construisent des séquences de sous-arbres emboîtés en supprimant de manière récursive les noeuds expliquant la moins grande déviance dans les données. La distinction entre arbre de classification et de régression provient de la variable réponse : si elle est catégorielle on utilise un arbre de classification (CTA), si elle est continue on utilise un arbre de régression (RTA). De nombreuses études ont montré que ces méthodes offraient une alternative aux régressions généralisées (GAM ou GLM), étant tout aussi précises et offrant des propriétés attrayantes comme la facilité de construction, la visualisation hiérarchique des variables sous une forme implicite et compréhensive, et la propriété d’inclure dans l’algorithme des interactions entre variables (Franklin 1998; Iverson & Prasad 1998; Vayssières *et al.* 2000). Leur inconvénient

majeur est la présence de discontinuités dans les nœuds et la faible adéquation des modèles pour des relations linéaires simples (Hastie & Tibshirani 1990; Moisen & Frescino 2002).

3.2.6 Méthodes d'apprentissage

Parmi les méthodes émergentes en biogéographie, les réseaux de neurones artificiels (ANN : Artificial Neural Networks) et les algorithmes génétiques (GA : Genetic Algorithm) sont les deux principaux exemples.

Les ANN apprennent à prédire sur la base de patrons identifiés dans des données d'apprentissage. Les ANN utilisent des combinaisons pondérées de manière optimale et non linéaire des variables originales pour parfaire la prédiction de la variable réponse. L'algorithme le plus communément utilisé est le '*back-propagation network*' (Openshaw & Openshaw 1997). Les données utilisées pour l'apprentissage traversent le réseau des centaines de fois jusqu'à ce que le réseau se stabilise. Pendant cette phase d'apprentissage, les nombreux paramètres du modèle sont estimés et optimisés. Les réseaux de neurones sont de plus en plus utilisés en biogéographie car ils offrent une bonne alternative aux méthodes linéaires trop contraignantes (Openshaw & Turner 2000; Pearson *et al.* 2002). Cependant, il est très difficile d'examiner les variables influençant la distribution des espèces. Ils sont donc peu explicatifs car ils ne permettent pas de tester des hypothèses sur la relation entre la distribution d'une espèce et celle d'une variable en particulier (Olden & Jackson 2002). Toutes ces imperfections font des ANN une 'boite noire' encore difficile d'accès pour beaucoup, et restent pour l'instant à un seul but prédictif et non explicatif comme peuvent l'être les GAM, GLM ou CART.

Les GA ont un processus d'optimisation basé sur une série de règles logiques (Openshaw & Openshaw 1997). Le GA le plus utilisé en biogéographie prédictive est GARP (Genetic Algorithm for Rule set Production, Stockwell 1999; Stockwell & Peters 1999; Stockwell *et al.* 2000), développé pour prédire la distribution d'espèces à partir de variables de milieu sur la base de cartes sous SIG (Système d'Information Géographique). Ce programme utilise un GA pour générer, tester et modifier des règles afin de prédire les distributions. Cela produit une série de règles qui peuvent prédire la présence ou l'absence d'une espèce pour des données d'apprentissage. Pour prédire la distribution d'espèces sur des sites non échantillonnés ou dans le future, GARP détermine quelle règle utiliser dans chaque situation sur la base de l'adéquation sur les données d'apprentissages. GARP fonctionne avec des données en présence-absence ou en présence seulement, mais dans ce cas des pseudo-absences sont créées (Stockwell & Peters 1999). Bien que très attractif, le manque d'interprétabilité des sorties, la procédure inaccessible et complexe, et le fait que peu d'études aient testé sa puissance

prédictive hormis celles des auteurs du programme, font que l'utilisation de cette méthode reste douteuse en biogéographie prédictive, où le fait de pouvoir interpréter les résultats par des mécanismes écologiques reste primordial.

3.2.7 Autres méthodes

D'autres méthodes existantes ne peuvent pas être classées dans les groupes précédents mais procurent de bonnes alternatives. On peut citer par exemple ENFA (Ecological Niche Factor Analysis, Hirzel *et al.* 2002a). Cette analyse, basée uniquement sur la présence des espèces, quantifie la niche écologique que l'espèce utilise en comparant sa distribution dans un espace défini par une ou plusieurs variables avec la distribution de toutes les cellules ('distribution globale') dans cet espace. ENFA permet ainsi de quantifier les propriétés de la niche des espèces en définissant une marginalité (comment la moyenne de l'espèce diffère de celle de la moyenne globale) et une tolérance (comment la variance d'habitats utilisés par l'espèce diffère de la variance globale). Cette analyse, testée par rapport à d'autres méthodes du type régression montre une bonne capacité de prédiction, mais présente le désavantage qu'il n'existe pas encore de méthode pour extrapoler la distribution d'une espèce dans une autre région ou dans le futur.

Après cette revue non exhaustive des différents modèles existants, il semble important d'examiner les effets de différents éléments d'un modèle sur la qualité des prédictions. Tout d'abord, avant de pouvoir prédire la distribution d'un organisme, il faut sélectionner un modèle. En 3.3, je présente deux analyses comparatives testant différents modèles en fonction de l'échelle et de la qualité des données. Un autre paramètre important est l'espèce modélisée et les variables qui lui sont associées. Je présente en 3.4, une étude montrant l'impact de la chorologie des espèces sur la fiabilité des prédictions. Enfin un autre paramètre fondamental, est la pertinence des variables explicatives. En 3.5, je teste la pertinence des variables d'utilisation des terres pour prédire les distributions à l'échelle européenne.

3.3 Choix des modèles

Chaque méthode en modélisation possède des spécificités propres qui font qu'elle est plus ou moins utile et fiable en fonction de l'échelle, du type de variable réponse ou bien encore de la qualité et de la pertinence des variables de milieu.

Suivent donc les conclusions de deux de mes analyses dont les buts étaient : 1- de comparer les méthodes les plus utilisées en biogéographie prédictive (GLM, GAM, CART) en fonction

Encadré I.1

Base de données :

Les trois bases de données sont sous la forme présence/absence (Tableau I.2). Les données d'espèces pour la Catalogne sont tirées de l'inventaire forestiers Catalan (IEFC, Gracia *et al.* 2000a, b, c, 2001a, b). Les données pour le Portugal sont tirées d'un projet mené par l'université d'Evora et rendus publique, et les données pour l'Europe sont tirées de l'Atlas Florae Europaeae (Jalas & Suominen 1972-1996).

Les variables de milieu utilisées pour ces trois analyses ont été choisies en fonction de leur disponibilité et sont considérées comme critiques pour la fonction physiologique des plantes et leur survie (Tableau I.2, Bartlein *et al.* 1986). Toutes les données espèces et milieu ont été traitées sous la forme de cartes au format grille ou point dans un système d'information géographique (SIG : Arcview 3.2.a). Les variables du milieu sont issues pour la Catalogne de l'Institut Cartographique Catalan, pour le Portugal de l'Atlas Environnemental (www.dga.pt) et l'Europe de données publiées (Legates & Wilmott 1990a, 1990b; Ahn & Tateishi 1994).

Modèles :

Dans cette étude, nous avons comparé 4 modèles :

- GLM 'simple' (McCullagh & Nelder 1989) : modèle logistique avec seulement des termes linéaires et interactions de premier ordre. Une procédure de sélection de variable a été utilisée pour sélectionner les variables les plus significatives en utilisant le critère AIC (Akaike Information Criterion, Akaike 1974).
- GLM 'complexe' (McCullagh & Nelder 1989) : modèle logistique avec termes linéaires, quadratiques et polynomiaux, et interactions de premier ordre uniquement sur les termes linéaires. La même procédure de sélection de variables a été utilisée.
- GAM (Hastie & Tibshirani 1990) : fonction de lissage du type cubic spline. Le degré de lissage a été automatiquement sélectionné par validation croisée et bornée à 4 pour chaque variable. Comme pour les deux GLM précédents, une procédure de sélection de variables sur critère AIC a été utilisée.
- CART (Breiman *et al.* 1984) : arbre de classification sur données binaires. 10 validations croisées successives ont été utilisées pour optimiser le compromis entre le nombre de branches de l'arbre et la déviance expliquée, pour être le plus parcimonieux possible tout en conservant une bonne fiabilité et précision.

Tous les modèles ont été calibrés sur une fraction aléatoire des données originales (70%), et leur puissance prédictive a été testée sur la fraction restante (30%).

Evaluation :

Pour évaluer le pouvoir prédictif des modèles, nous avons choisi la méthode Receiver Operating Characteristic curve (courbe ROC, Hanley & McNeil 1982) qui évalue graphiquement la fraction des présences correctement prédites contre la fraction des présences non correctement prédites (Figure I.4-A). La courbe ROC indique le pourcentage de décisions correctes dans des comparaisons par paire (Swets 1988), qui dans un contexte de prédictions d'espèce revient à estimer la probabilité de classer correctement une présence ou une absence. Si toutes les prédictions étaient faites au hasard, la relation serait une droite de pente 0,5 et d'intercepte 0 (1^{ère} bissectrice). L'estimation de l'aire comprise entre la courbe ROC (AUC : aire au dessous de la courbe) et la première bissectrice représente cette qualité de discrimination (Figure I.4-A). L'AUC varie entre 0,5 et 1. Une AUC de 1, indique une discrimination parfaite alors qu'une AUC de 0,5 est équivalente à des prédictions aléatoires.

La règle subjective suivante a été employée concernant l'interprétation des valeurs de AUC (Swets 1988) : 0,90-1 = excellent discrimination; 0,80-0,90 = bonne; 0,70-0,80 = moyenne; 0,60-0,70 = pauvre; 0,50-0,60 = nulle (Swets 1988).

La différence entre les AUC de deux modèles différents produit une mesure comparative de discrimination de ces modèles lorsqu'ils sont appliqués à des données d'évaluation. Dans notre étude, la différence entre deux AUC (AUC1 et AUC2) a été calculée selon le test ratio (Hanley & McNeil 1982)

$$Z = \frac{AUC1 - AUC2}{\sqrt{Se^2_{AUC1} + Se^2_{AUC2} - 2r * Se_{AUC1} Se_{AUC2}}}$$

Se est l'erreur standard et *r* est la corrélation entre les deux aires en dessous de la courbe ROC.

Enfin, une fois l'étape d'évaluation réalisée, un seuil maximisant le pourcentage de présences et d'absences correctement prédites (Pearce & Ferrier 2000) a été utilisé pour cartographier les distributions simulées sous la forme de présence ou d'absence et non sous la forme de probabilité de présence (Figure I.4-B).

de la complexification des relations distribution-milieu liée à l'échelle spatiale, 2- de comparer deux types de méthodes (GLM et ENFA) en fonction du type de données (présence/absence *versus* présence seulement).

3.3.1 Complexification des relations distribution-milieu liée à l'échelle

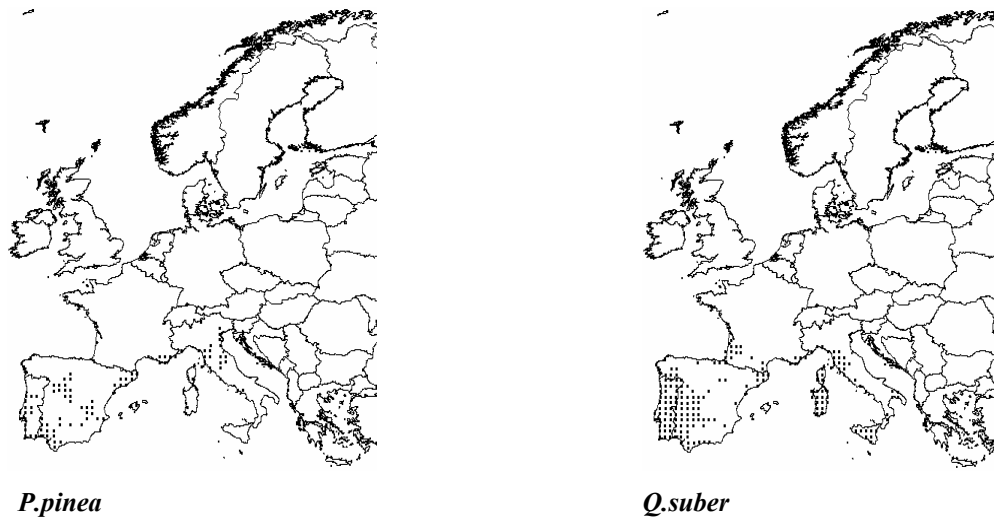
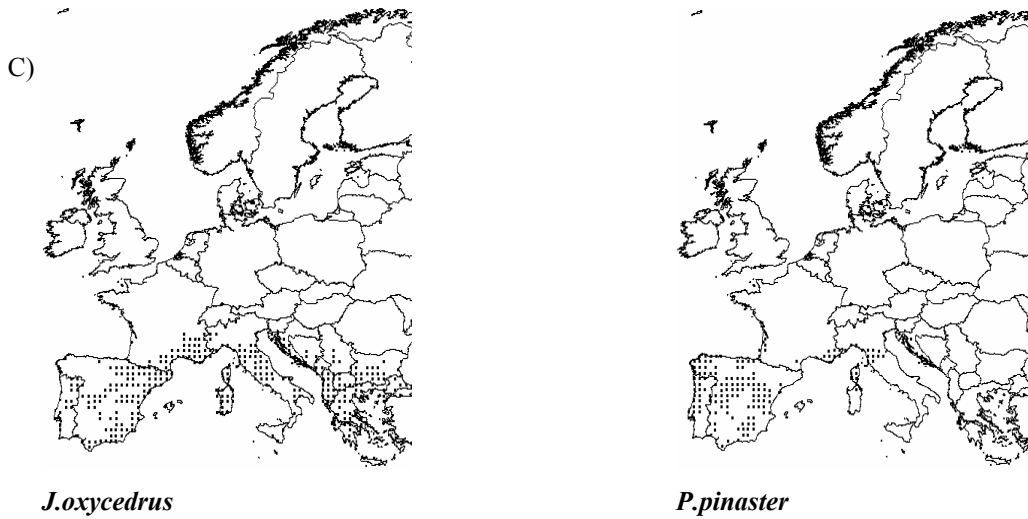
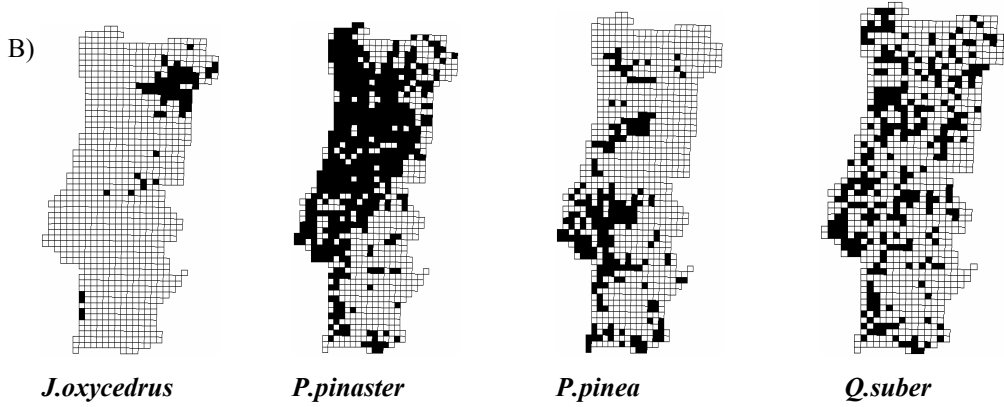
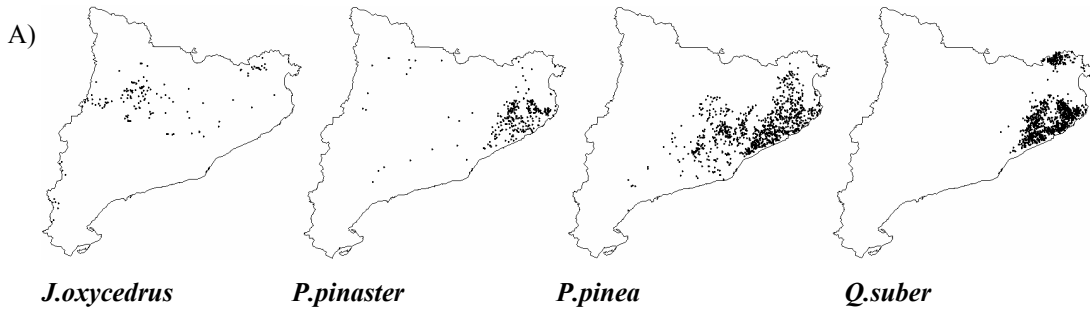
Basé sur Annexes 1.1 et 1.2 :

- Thuiller, W., Araújo, M.B. & Lavorel, S. (2003). Generalized Models vs Classification Tree Analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, 14 : 669-680.
- Thuiller, W., Howell, C. A., Midgley, G. F. & Rouget, M. (En préparation). Evaluating species distribution models: methodologies and conservation implications of prevalence and probability thresholds.

Parmi toutes les méthodes présentées précédemment, les outils de régression (GLM et GAM) et les arbres de classification (CTA) sont les plus utilisés pour décrire et prédire la distribution des espèces. Différentes études ont comparé la puissance prédictive de ces trois types d'approches. Par exemple, Vayssières *et al.* (2000) comparent la puissance prédictive de CTA et GLM sur 8 espèces de chêne en Californie. Les auteurs montrent que les modèles CTA sont significativement meilleurs pour quatre des six espèces, et aussi bon que les GLM pour les deux autres. Ils ajoutent que les CTA possèdent une capacité supérieure aux GLM pour estimer les degrés d'interactions entre variables, et offrent une forme intuitive de modélisation. De même, Franklin (1998) explore la puissance prédictive de CTA, GAM et GLM pour prédire la distribution de vingt arbustes dans le sud-ouest de la Californie. L'auteur conclut que bien que les trois méthodes aient des puissances prédictives similaires, les GAM sont très utiles pour explorer les fonctions réponses, les GLM permettent à ces fonctions d'être paramétrées et de tester leur significativité, et les arbres de classifications, bien que difficiles à interpréter, ont la plus faible erreur de prédiction (3-5%). Ces deux analyses comparatives sont très utiles, cependant, les conclusions restent peu généralisables car la notion d'échelle ou de résolution des données n'est pas abordée. En effet, les méthodes comparées ne reposent pas sur les mêmes algorithmes et les mêmes propriétés, ce qui laisse penser qu'elles n'ont pas la même aptitude à modéliser la distribution d'une espèce en fonction de la complexification des relations espèce-milieu sous-jacente à un changement d'échelle.

Pour tester la différence de puissance prédictive entre ces trois méthodes, j'ai réalisé une analyse sur 4 espèces ligneuses méditerranéennes : le genévrier oxycèdre (*Juniperus oxycedrus*), le chêne liège (*Quercus suber*), le pin maritime (*Pinus Pinaster*) et le pin pignon (*Pinus pinea*). Des modèles ont été utilisés pour analyser les déterminants climatiques de ces quatre espèces et pour prédire leurs distributions actuelles à 3 échelles et résolutions

Figure I.3 : Distributions observées des espèces sélectionnées. A) Catalogne, B) Portugal et C) Europe.



décroissantes : Europe, Portugal et Catalogne. Le choix des espèces a été fortement influencé par la disponibilité des données d'espèces (Encadré I.1).

3.3.1.1 Biogéographie des espèces :

Le genévrier oxycèdre est un arbre dioïque atteignant 14 m, mais est souvent rencontré à l'état d'arbuste. Sa répartition en Europe est principalement méditerranéenne entre 0 et 1400 m (Figure I.3). Le pin maritime est, quant à lui, un arbre (30-40 m) de l'ouest du bassin Méditerranéen, présent dans le nord-ouest de l'Afrique, le Portugal ou l'Espagne. En Europe continentale, il se situe entre les altitudes 0 et 400 m, bien qu'il puisse pousser en Corse ou au Maroc aux alentours de 2000 m. C'est une espèce largement plantée et maintenant naturalisée en Afrique du Sud (Richardson 1998). Le pin pignon est un arbre plus petit (12-25 m) typique des paysages méditerranéens. Il a été si largement planté tout autour du bassin Méditerranéen, qu'il est aujourd'hui impossible de déterminer sa répartition naturelle (Baquero & Telleria 2001). Cette espèce a été récemment naturalisée en Afrique du Sud et autres régions du monde à climat méditerranéen (Richardson 1998).

3.3.1.2 Effets de l'échelle et de la complexification des relations espèce-milieu :

La puissance prédictive des modèles a montré différents patrons relatifs à l'échelle d'étude. Quelles que soient l'échelle et la résolution associée, la fiabilité des GAM ne semble pas être perturbée car ils sont non paramétriques et peuvent faire l'approximation des différentes formes de réponses (Guisan *et al.* 2002; Oksanen & Minchin 2002). Ceci est d'autant plus bénéfique que les courbes de réponse des espèces aux facteurs de milieu sont susceptibles de varier avec l'échelle, particulièrement pour les espèces à vaste répartition (ex : *Quercus suber* dans cette étude). En Catalogne, *Q. suber* a une répartition qui s'étend au delà de la zone d'étude et donc la forme de la courbe de réponse à un fort gradient climatique est censée être monotone car elle reflète seulement partiellement la niche réalisée complète de l'espèce. A cette échelle, le GLM incluant seulement des termes linéaires est donc suffisant pour prédire la distribution de l'espèce (Figure I.5-A). Inversement, à échelle européenne, la forme de la courbe de réponse décrit une plus large proportion de la niche réalisée. Comme montré pour *Q. suber* (Figure I.5-B), la forme de la courbe de réponse est Gaussienne ou asymétrique, mais pas linéaire (Huisman *et al.* 1993; Austin & Gaywood 1994; Oksanen & Minchin 2002). Les GLM incluant des fonctions plus complexes, telles que les termes quadratiques ou polynomiaux sont plus appropriés pour estimer la forme de la courbe de réponse à mesure que

Figure I.4: A) Courbe ROC représentant le pourcentage de présences correctement prédites (vraies positives) contre le % d'absence incorrectement prédites (fausses positives) calculés pour différents seuils de probabilité. B) % de présences et d'absences correctement prédites en fonction du seuil de probabilité. Le point de croisement des deux courbes correspond au seuil maximisant les deux proportions.

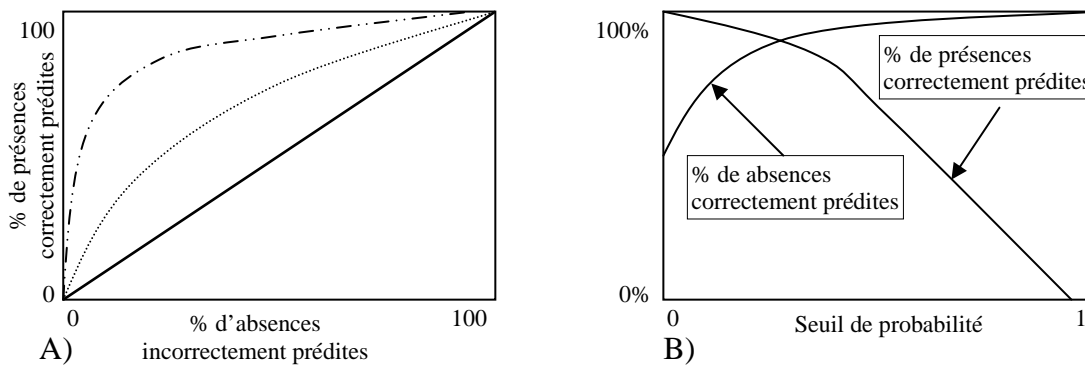


Figure I.5 : Courbe de réponse de l'espèce *Quercus suber* à la température du mois le plus froid en Catalogne et du mois le plus chaud en Europe, simulée par GLM

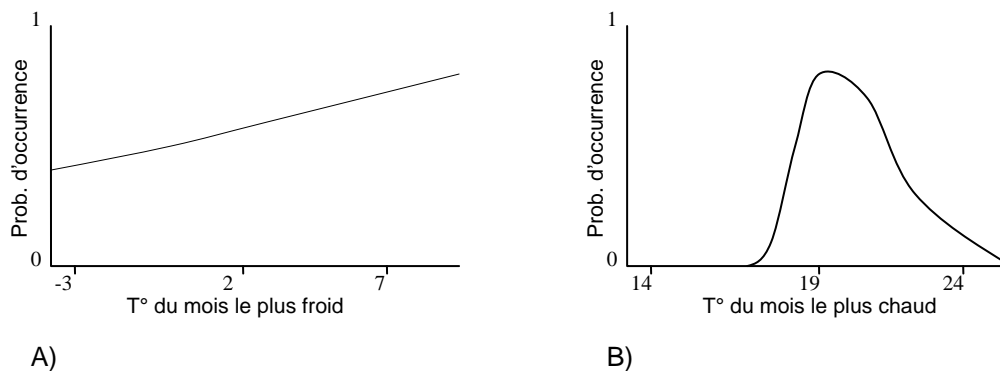


Tableau I.2 : Récapitulatif des surfaces, des résolutions et des variables de milieu pour les trois zones d'études.

	Catalogne	Portugal	Europe
Surface (km ²)	32 113	92 082	5.722.500
Résolution (km)	0.18x0.18	10x10	50x50
Nombre de points d'échantillonnage	10639	993	2289
Moyenne des précipitations annuelles	X	X	X
Moyenne des précipitations saisonnières	X		
Moyenne des températures annuelles	X	X	
Moyenne des températures du mois le plus froid	X		X
Moyenne des températures du mois le plus chaud	X		X
Moyenne annuelle des températures maximales	X		
Moyenne des températures maximales du mois le plus chaud	X		
Moyenne annuelle des températures minimales	X		
Moyenne des températures minimales du mois le plus froid	X		
Moyenne des radiations annuelles	X		
Moyenne annuelle d'évapotranspiration potentielle			X
Evapotranspiration du mois le plus froid			X
Evapotranspiration du mois le plus chaud			X
Altitude	X	X	X
Pente	X	X	
Géologie	X		

la proportion de la niche réalisée des espèces incluse dans l'analyse augmente. Les CTA, quand à eux, n'ont jamais une puissance prédictive supérieure aux méthodes généralisées. A l'échelle la plus fine, les CTA ont la puissance prédictive la plus faible. En comparaison à des études précédentes (Vayssières *et al.* 2000; Rouget *et al.* 2001b), ceci peut être attribué à l'absence d'effets interactifs parmi les variables de milieu (Hastie & Tibshirani 1990). L'augmentation de la puissance prédictive de CTA avec l'augmentation de l'échelle, pourrait en effet refléter des relations plus complexes parmi les variables. Néanmoins, CTA reste une méthode qui rend possible la description des relations espèces-milieu, et de les résumer sous la forme d'un arbre facilement lisible pour les non-initiés.

3.3.1.3 Facteurs influençant l'adéquation des modèles

De nombreux facteurs peuvent influencer les modèles et leurs qualités d'ajustements. Premièrement, cette étude a montré que les espèces qui ont une distribution compacte ou agrégée sont plus facilement modélisables que les espèces à distribution plus dispersée. Ce phénomène a déjà été décrit (Araújo & Williams 2000) et rejoint le précédent point concernant l'amplitude des gradients capturés par les modèles. En effet, à échelle fine, une espèce avec une distribution compacte a probablement une part plus forte de sa distribution comprise dans la zone d'étude, augmentant l'aptitude des modèles à capturer sa niche réalisée.

Deuxièmement, l'hypothèse d'équilibre sous-jacente aux méthodes statiques peut être mise en doute. Des distributions non à l'équilibre, pouvant être le résultat de facteurs historiques, interactions biotiques, perturbations (feu) ou impacts de l'homme, peuvent rendre les prédictions risquées (Bolliger *et al.* 2000). Cependant, l'impact de la gestion des forêts ou les feux fréquents qui influencent la distribution des arbres, reste difficile à évaluer (Duckworth *et al.* 2000a) bien que Rouget *et al.* (2001a) ait montré qu'en Catalogne le feu avait peu d'effets sur les distributions à l'équilibre des espèces de pins. Par exemple, *P.pinaster* et *P.pinea* ont été intensivement plantés et leurs distributions pourraient bien ne pas être à l'équilibre avec les conditions de milieu actuelles, expliquant pourquoi les modèles sont peu prédictifs pour ces espèces à fine échelle (Rouget *et al.* 2001). Cependant, Bolliger *et al.* (2000) soutient qu'à échelle régionale, les méthodes statiques sont robustes dans la prédiction des espèces, même si elles ne sont pas totalement à l'équilibre. Ces auteurs montrent aussi qu'à cette échelle les modèles statiques sont aussi

Encadré I.2

Données espèces : Nous avons utilisé un échantillon d'espèces de l'atlas des oiseaux nicheurs de la Catalogne (Catalan Breeding Birds Atlas CBBA, 1999-2002). 1500 cellules de 1x1 km (6% de la surface totale) ont été sélectionnées pour la conduite de surveillance standardisée de présence des espèces durant les saisons de nidification 1999-2001. Dans ces cellules, la présence relevée d'une espèce est une donnée quasi certaine tandis que la non observation d'une espèce ne veut pas nécessairement dire que l'espèce est absente et que l'habitat n'est pas favorable. Nous avons travaillé sur ces données en sélectionnant 30 espèces avec un nombre d'occurrence supérieur à 15.

Variables de milieu : Nous avons utilisé des données bioclimatiques et des données d'utilisation des terres disponibles en Catalogne à 1x1 km. Les données bioclimatiques sont les mêmes que celles utilisées dans l'étude précédente. Les données d'utilisation des terres sont des données de recouvrement (ex : forêts de chênes lièges ou de pins), d'occupation du sol par l'agriculture et de degré d'urbanisation.

Modèles : La méthode utilisant des données en présence-absence que nous avons utilisée est un GLM avec des termes linéaires et quadratiques, et un processus de sélections de variables similaires à l'étude précédente.

La méthode utilisant les données de présence seulement est ENFA (Hirzel *et al.* 2002).

Evaluation : Utilisation des courbes ROC et AUC associées (Cf. Encadré I.1).

fiables que les modèles dynamiques incluant des paramètres physiologiques et les traits d'histoire de vie des espèces forestières.

3.3.2 Effets de l'échantillonnage : données présence-absence *versus* présence seulement

Basé sur Annexe 1.3 :

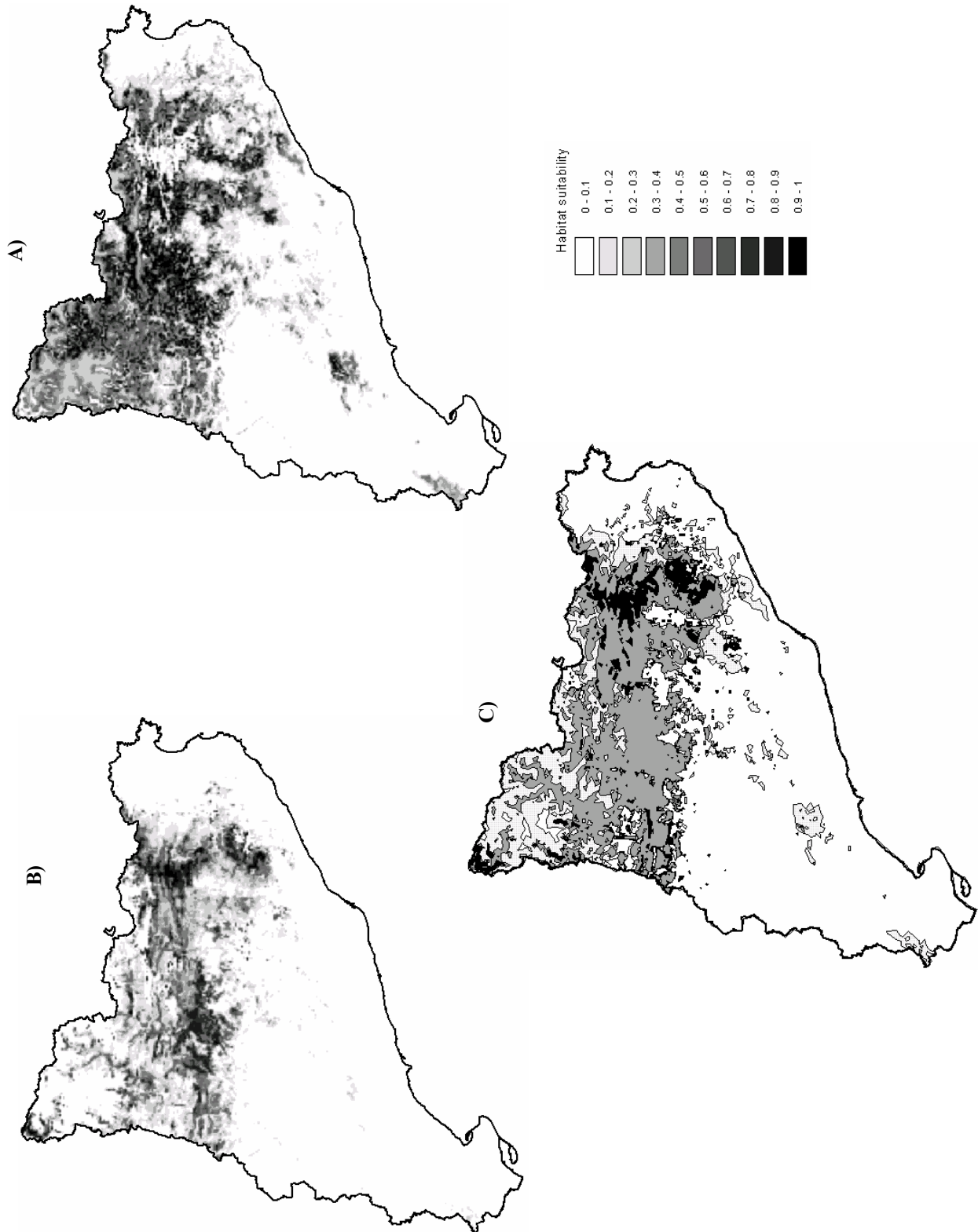
Brotons L., Thuiller W., Araújo M.B. & Hirzel A.H. (2004) Presence-absence vs presence-only modelling methods for predicting bird habitat suitability, *Ecography*, sous presse

Parmi les modèles basés sur les niches certains permettent d'utiliser des données de présence-absence et d'autres seulement des données de présence. Même si la puissance des modèles et la qualité des variables prédictives sont des facteurs primordiaux pour la prédiction des distributions d'espèces, la forme des données d'espèces géographiquement référencées est un paramètre très important. Parce qu'ils sont mobiles, les données sur les animaux sont souvent des observations ponctuelles où les données de présences sont de vraies présences mais où les données d'absences peuvent être simplement le fait de la non observation. Très fréquemment les atlas d'oiseaux (Hagemeijer & Blair 1997) ou de mammifères (Mitchell-Jones *et al.* 1999) recensent les occurrences d'un grand nombre d'observations et de données de muséums (Peterson *et al.* 2002). Sur ces grilles de données, la présence d'une espèce donnée est assez fiable tandis que la fiabilité des absences, quand elles sont recensées, reste problématique. Certaines méthodes statiques utilisent uniquement des données de présences et sont donc particulièrement attractives pour analyser de tels types de données (Tableau I.1). Toutefois, il semble naturel de se demander si ces méthodes sont aussi fiables que celles utilisant des données en présence-absence.

Nous avons donc utilisé un atlas d'oiseaux nicheurs en Catalogne pour analyser deux types de modèles : Un modèle basé uniquement sur des données de présence (Ecological Niche Factor Analysis, ENFA : Hirzel *et al.* 2002a) et un autre basé sur des présence-absence (GLM). Le choix de ces deux types de méthodes est justifié par le fait que ENFA est une méthode spécifiquement créée pour traiter ce genre de données, et que GLM est la méthode la plus fréquemment utilisée par les écologues et ornithologues. Ces deux modèles ont été utilisés pour prédire la distribution de 30 espèces d'oiseaux nicheurs en Catalogne et pour comparer leur puissance prédictive (Encadré I.2).

Les résultats de cette analyse ont montré que les modèles GLM avaient un meilleur pouvoir prédictif que ENFA. Ce résultat soutient l'idée que les espèces de forêt analysées dans

Figure I.6 : Distribution de la Sittelle torchepot (*Sitta Europaea*) prédite par GLM (A) et ENFA (B), et les différences entre les deux prédictions (C). La carte de divergence a été créée en croisant les deux cartes provenant des deux approches, après les avoir transformées en présence/absence en maximisant le % de présence et d'absence correctement prédites (Cf. Encadré I.1). Les cellules noires montrent les zones où ENFA prédit l'absence de l'espèce et GLM la présence, tandis que les cellules gris clair montrent les zones où ENFA prédit la présence et GLM l'absence de la sittelle. Les cellules gris foncé indiquent la coïncidence entre les deux prédictions.



cette étude utilisent les habitats disponibles proportionnellement à leur adéquation au milieu, faisant que les données d'absence sont fiables et utiles pour améliorer la calibration du modèle (Hirzel *et al.* 2001). En sous-pondérant les occurrences des espèces dans des habitats de mauvaise qualité, les données d'absence aident à identifier ces zones qui pourraient être classées comme de bons habitats si les données de présence seulement étaient utilisées. Par exemple, dans le cas de la Sittelle torchepot (*Sitta Europaea*), les zones de forêts de décidus qui couvrent une surface limitée à l'intérieur de la zone d'étude, sont classées comme faiblement adéquates par ENFA (Figure I.6). En effet, cette espèce a un faible nombre d'occurrences dans ces zones comparé au grand nombre d'occurrences dans les habitats plus abondants. Dans de tels cas, la disponibilité des données d'absence joue un rôle primordial pour évaluer correctement l'adéquation relative des différents habitats. Récemment, Zaniewski *et al.* (2002) a montré que bien que les méthodes basées sur présence-absence étaient plus discriminantes que ENFA appliquée à l'espèce, elles apparaissaient moins performantes pour identifier des zones à fort caractère de conservation d'un point de vue multi-spécifique. Dans ce cas, si l'objectif de l'étude est de protéger des espèces rares ou potentiellement en danger, la surestimation des zones ayant potentiellement une biodiversité élevée peut être préférable à la sous-estimation (Zaniewski *et al.* 2002).

A échelle assez large, beaucoup d'espèces et spécialement les espèces mobiles sont proches de l'équilibre avec les conditions de milieu en raison de leur rapidité de réponse à un changement de milieu et des mécanismes de sélection d'habitat (Chamberlain & Fuller 1999). Dans ces cas, les absences reflètent probablement des conditions d'habitats inappropriées et permettent d'augmenter la qualité prédictive des modèles. En conclusion, les données de présence/absence sont plus appropriées que les présences seulement pour construire des modèles d'habitats favorables (*'Habitat Suitability Models'*) et de distribution d'espèces tels que les collections de données intensives des atlas d'oiseaux. Dans ces approches, la distribution spatiale déterminée par la niche réalisée de l'espèce est très utile (Pulliam 2000). L'interprétation écologique des méthodologies doit guider le choix final parmi plusieurs alternatives. Zaniewski *et al.* (2002) soutient que les méthodes utilisant seulement des données de présence prédisent les distributions potentielles des espèces et s'approchent donc de leur niche fondamentale, alors que les modèles basés sur présence/absence reflètent plus vraisemblablement les distributions naturelles actuelles dérivées des niches réalisées. Cependant, le but de ces deux approches est de prédire la distribution des espèces à partir d'échantillonnages de la distribution réelle des espèces, et donc procurent plutôt différentes

Encadré I.3

Données espèces : L'inventaire forestier de la Catalogne (IEFC) est une base de données de 10638 points d'échantillonnage aléatoire (rayon de 10 m) dans les forêts de Catalogne. Cet échantillonnage a été réalisé par le CREAM de Barcelone. L'IEFC regroupe 95 espèces ligneuses, incluant les genres les plus abondants tels que *Pinus* (7 esp.) et *Quercus* (12 esp.).

Pour cette étude nous avons sélectionné 24 essences représentatives de la diversité compositionnelle de la Catalogne. Ce groupe d'espèces contient 12 genres : *Quercus* (4 esp.), *Pinus* (6 esp.), *Acer* (2 esp.), *Juniperus* (2 esp.), *Populus* (2 esp.), *Abies*, *Castanea*, *Corylus*, *Fagus*, *Salix*, *Sorbus* et *Ulmus* (1 esp.). Ce groupe est aussi représentatif de la variété existante des patrons phytogéographiques, incluant des espèces boréo-alpines (2 esp.), euro-sibériennes (10 esp.), sub-méditerranéennes (6 esp.) et méditerranéennes (6 esp.). Ces patrons phytogéographiques correspondent aux espèces dites au centre (méditerranéennes et sub-med) et celles aux bornes (boréo-alpines et euro-sibériennes) de leur distribution biogéographiques (Quézel 1985).

Variables de milieu : les mêmes que l'Encadré 1, plus une variable géologique catégorielle (Tableau I.3). Chaque variable (180x180 m de résolution) a été projetée sur les points d'échantillonnage pour avoir en chaque point, la présence ou l'absence d'une espèce ainsi que les variables de milieu (Tableau I.3).

Modèles : Etant donné le but de l'exercice qui est de décrire les principaux déterminants climatiques et topographiques des espèces et chorotypes associés, nous avons utilisé des arbres de classifications (CTA). Cet outil permet d'identifier rapidement de manière hiérarchique les relations entre la distribution des espèces et les variables sélectionnées. Les CTA ont été calibrés sur un échantillonnage aléatoire des données initiales (70%).

L'évaluation des modèles s'est faite avec la courbe ROC comme explicité précédemment sur les données d'évaluation (30% des données initiales). Nous avons également utilisé les données de milieu originales (180x180 m) pour projeter les habitats potentiels des essences sélectionnées.

Tableau I.3 : Liste des variables de milieu utilisées

Variables de milieu	Abréviation
Somme des précipitations annuelles	MAP
Somme des précipitations en été	SumPrec
Somme des précipitations en hiver	WintPrec
Somme des précipitations en automne	AutPrec
Somme des précipitations au printemps	SprPrec
Moyenne des températures annuelles	MAT
Moyenne des températures du mois le plus froid	MTC
Moyenne des températures du mois le plus chaud	MTW
Moyenne des températures maximales	MAMiT
Moyenne des températures maximales du mois le plus chaud	MMTC
Moyenne des températures minimales	MAMxT
Moyenne des températures minimales du mois le plus froid	MMTW
Moyenne des radiations annuelles	MAR
Pente	Slope
Géologie (variable catégorielle)	Geol
catégorie 1: Roches siliceuses	
catégorie 2: Roches carbonatées	
catégorie 3: Calcaire	
catégorie 4: Sédiments calcaires	
catégorie 5: Sédiments siliceux	

estimations de la niche réalisée des espèces (Loehle & LeBlanc 1996). Comme les méthodes utilisant seulement des données de présences ne tiennent pas compte des zones où l'espèce peut être absente, elles sont tout simplement moins conservatives en estimant la niche réalisée de l'espèce. Autrement dit, cette approche pourrait être utile pour capturer la niche réalisée des espèces qui sont loin de l'équilibre parce que n'ayant pas encore atteint tous les habitats favorables pour elles (Hirzel *et al.* 2001).

3.4 Choix des espèces : Impacts de la chorologie

Basé sur Annexe 1.4 :

Thuiller, W., Vaydera, J., Pino, J., Sabaté, S., Lavorel, S. & Gracia, C. (2003). Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology & Biogeography*, 12, 313-325.

Dans cette étude nous avons examiné si des espèces de chorotypes différents dans une région particulière répondaient aux mêmes variables de milieu et si les modèles étaient capables de prédire leur distribution actuelle et potentielle avec la même efficacité. Dans cette optique, la région de la Catalogne offre une zone particulièrement intéressante d'un point de vue biogéographique, compte tenu des forts gradients altitudinaux et des influences climatiques qui s'y affrontent. De plus, cette région offre une grande diversité floristique, et particulièrement au niveau du peuplement forestier où un grand nombre d'essences se côtoient. Une collaboration avec des chercheurs (Joan Pino, Santi Sabate, Jordi Vayredas et Carles Gracia) du CREAM (Centre de Recherche en Ecologie et Applications Forestières) nous a permis d'avoir accès à l'inventaire forestier effectué dans cette région (Encadré I.3).

3.4.1 Déterminants climatiques et géologiques des distributions d'espèces forestières représentatives de Catalogne

La précision des modèles selon la courbe ROC augmente des espèces eurosibériennes (50% des espèces classées dans les classes moyennes et hautes précisions), aux espèces méditerranéennes (83%) (Tableau I.4). Il y a donc une tendance d'augmentation de la qualité des prédictions des espèces azonales (espèces aux bornes de leur aire de répartition) aux espèces zonales (espèces au centre de leur aire de répartition).

En ce qui concerne les variables climatiques, cinq d'entre elles sont fréquemment sélectionnées par les CTA (Tableau I.5) : trois variables de précipitation (annuelle, été et hiver) et deux de températures (maximum annuelle, mois le plus froid). Parmi les variables topographiques, la géologie est celle qui est le plus souvent sélectionnée. Ces variables ont des

Tableau I.4 : Exactitude des prédictions pour les 24 espèces sélectionnées. C=Chorologie. B=Boréoalpine, E=Eurosibérienne, M=Méditerranéenne, SM=Subméditerranéenne. *n* est le nombre d'occurrences de chaque espèce. Pres et Abs correspondent respectivement au % moyen de présence et d'absence correctement prédites.

Précision	Espèces	C	<i>n</i>	Evaluation		
				Pres	Abs	AUC
Haute	<i>Pinus uncinata</i>	B	787	90.6	94.7	0.979
	<i>Quercus suber</i>	M	1034	92.0	93.6	0.956
	<i>Fagus sylvatica</i>	E	567	85.8	85.4	0.941
	<i>Pinus halepensis</i>	M	2712	87.1	86.8	0.941
	<i>Abies alba</i>	B	229	93.0	89.6	0.937
	<i>Castanea sativa</i>	SM	246	90.0	82.5	0.932
Bonne	<i>Pinus pinea</i>	M	997	88.8	74.8	0.891
	<i>Pinus nigra</i>	SM	1971	89.6	75.5	0.890
	<i>Populus tremula</i>	E	151	75.7	84.9	0.886
	<i>Pinus sylvestris</i>	SM	3217	75.8	78.7	0.875
	<i>Acer opalus</i>	SM	240	78.7	77.8	0.835
	<i>Sorbus aria</i>	E	140	77.1	80.0	0.832
	<i>Quercus petraea</i>	E	262	81.4	71.2	0.817
	<i>Juniperus oxycedrus</i>	M	146	65.8	86.3	0.813
	<i>Pinus pinaster</i>	M	302	78.5	81.2	0.81
	<i>Salix caprea</i>	E	91	79.3	81.0	0.808
Moyenne	<i>Quercus humilis</i>	SM	1847	79.0	68.7	0.797
	<i>Populus nigra</i>	E	102	59.4	86.5	0.796
	<i>Acer monspessulanum</i>	SM	157	64.9	72.1	0.785
	<i>Quercus ilex</i>	M	4070	74.5	68.7	0.779
	<i>Ilex aquifolium</i>	E	102	76.2	71.6	0.765
	<i>Corylus avenalla</i>	E	182	73.7	63.5	0.752
Faible	<i>Juniperus communis</i>	E	212	68.9	59.1	0.646
	<i>Ulmus minor</i>	E	115	63.9	65.2	0.645

poids différents selon les espèces et les genres. Par exemple, les modèles sélectionnent la précipitation en hiver comme le déterminant climatique majeur des espèces de pins, sauf pour le pin d'Alep (*Pinus halepensis*, Figure I.7-A), qui semble être principalement affecté par la précipitation annuelle. La température du mois le plus froid semble être un facteur limitant pour le pin pignon (*Pinus pinea*), le pin maritime (*Pinus pinaster*) et le pin d'Alep.

Le Tableau I.5 révèle aussi des patrons à l'intérieur des classes chorologiques. Les espèces méditerranéennes sont distribuées principalement suivant la température et ensuite suivant la précipitation. Inversement, les espèces sub-méditerranéennes et eurosibériennes (ex : *Fagus sylvatica*, Figure I.7-B) semblent être influencées principalement par la précipitation et ensuite par la température. Il est plus difficile de trouver un patron pour les espèces boréoalpines, étant limitées en nombre (2 espèces, ex. *Pinus uncinata*, Figure I.7-C).

Cette étude permet de confirmer l'importance de la précipitation annuelle et des températures extrêmes pour expliquer et prédire les patrons régionaux de la distribution des espèces méditerranéennes (Barbéro *et al.* 1998; Rouget *et al.* 2001a). Par exemple, *Pinus halepensis* est supposé se trouver des habitats semi-arides aux habitats semi-humides (Barbéro *et al.* 1998). Notre modèle corrobore ces résultats, montrant que la distribution de ce pin est très fortement liée à des précipitations annuelles moyennes et est restreint à des températures douces en hiver (Figure I.7-A). Cette espèce est bien adaptée aux conditions méditerranéennes et la zone d'étude est située au centre de sa répartition européenne, faisant que la prédiction de son habitat potentiel est assez juste. A l'inverse, les facteurs du paysage (ex : convexité, profondeur de la nappe phréatique) pourraient être plus importants que les moyennes climatiques pour les espèces aux bornes de leur aire de répartition, comme les espèces eurosibériennes qui sont restreintes à des conditions topographiques et climatiques locales (marginales pour la région d'étude, donc définie par des conditions exceptionnelles ayant un effet très localisé). Ceci est particulièrement vrai pour des espèces de ripisyles telles que *Populus nigra* et *Ulmus minor*. L'échec relatif des modèles pour ces espèces pourrait être gommé en travaillant à échelle plus fine et avec des variables topographiques et géomorphologiques (Reed *et al.* 1993; Bridge & Johnson 2000). Cependant, les modèles peuvent aussi engendrer dans certains cas de bonnes prédictions pour des espèces en marge de distribution comme *Fagus sylvatica* (Figure I.7-B). Cette espèce continentale se situe principalement dans le centre de l'Europe, la Catalogne étant sa limite sud. Prentice, Bartlein *et al.* (1991) et Huntley, Berry *et al.* (1995) ont montré que cette espèce se situait dans des régions avec des précipitations annuelles proche de 1200 mm et était absente des zones où la température moyenne du mois le plus froid était inférieure à -3°C. Notre étude suggère les

Tableau I.5 : Liste des 4 premières variables de milieu sélectionnées par les CTA pour chaque espèce (en terme de déviance). Les 5 dernières lignes représentent l'importance relative de chaque variable sur les 24 espèces. Pour chaque espèce, un index a été attribué suivant l'ordre de sélection par le modèle. La note de 4 pour la 1^{ère} variables, 3 pour la 2^{nde}, 2 pour la 3^{ème} et 1 pour la 4^{ème}. La chorologie des espèces a été assignée suivant Bolos *et. al.* (1990). Voir Tableau I.3 pour l'abréviation des variables.

	MAP	WinPrec	SprPrec	SumPrec	AutPrec	MAT	MAMiT	MTC	MMTC	MAMxT	MTW	MMTW	MAR	Slope	Geol
Boréoalpine															
<i>Abies alba</i>		2		3							1		4		
<i>Pinus uncinata</i>		3					2			1					
Eurosibérienne															
<i>Corylus avellana</i>	1		2			3									4
<i>Fagus sylvatica</i>	1	3		4						2					
<i>Ilex aquifolium</i>	1							3			4		2		
<i>Juniperus communis</i>		3		1	4		2								
<i>Populus nigra</i>		4										3		1	2
<i>Populus tremula</i>	1		2							3			4		
<i>Quercus petraea</i>			1	3						2		4			
<i>Salix caprea</i>				2		1			4						3
<i>Sorbus aria</i>					3					4	1			2	
<i>Ulmus minor</i>				2				1		3			4		
Subméditerranéenne															
<i>Acer monspessulanum</i>	4				1	2							3		
<i>Acer opalus</i>			3						4		1				2
<i>Castanea sativa</i>	3								2				4		1
<i>Pinus nigra</i>		1	3	4		2									
<i>Pinus sylvestris</i>	3	4		1											
<i>Quercus humilis</i>				1	4				2	3					
Méditerranéenne															
<i>Juniperus oxycedrus</i>					1			2	4			3			
<i>Pinus halepensis</i>	1							2	3				4		
<i>Pinus pinaster</i>		2			4		3	1							
<i>Pinus pinea</i>	3			4			2	1							
<i>Quercus ilex</i>		2		3						1					4
<i>Quercus suber</i>		4						1	2						3
1 ^{ère} position	5	1	1	3	2	1	0	4	0	2	3	0	0	1	1
2 ^{ème} position	0	3	2	2	0	2	4	2	3	2	1	0	1	1	2
3 ^{ème} position	3	3	2	3	1	1	1	1	1	3	0	2	1	0	2
4 ^{ème} position	1	3	0	2	2	0	0	0	3	4	4	1	4	0	2
Importance des variables	27	22	14	26	12	12	14	24	14	24	19	5	9	7	16

mêmes restrictions dues aux limites de températures du mois le plus froid et des précipitations annuelles avec des valeurs similaires à celles déterminées par ces auteurs.

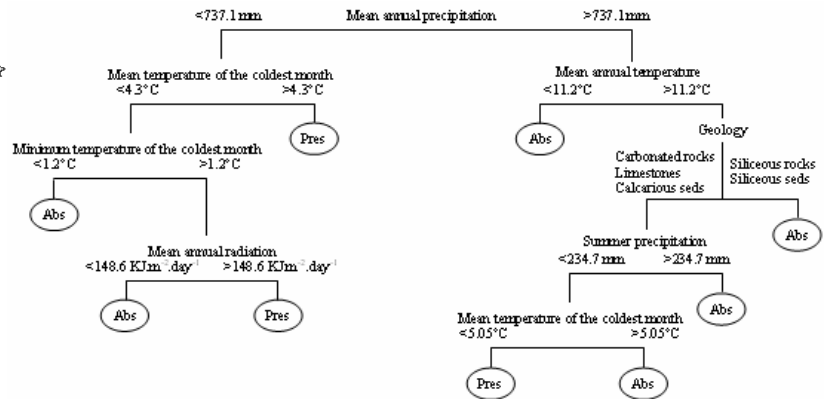
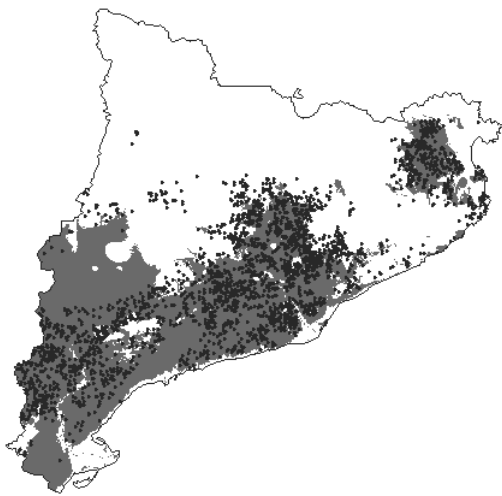
3.4.2 Biogéographie et chorologie

Logiquement, les espèces les mieux modélisées sont celles poussant dans la région d'étude dans une grande proportion de leurs habitats potentiels pouvant être considérées comme proche de l'équilibre avec leur milieu. C'est le cas du pin d'Alep (*Pinus halepensis*), qui grâce à sa forte capacité de colonisation est seulement absent de sa distribution potentielle dans les zones intérieures arides, historiquement vouées à l'agriculture, et dans l'extrême sud-est dû certainement aux feux fréquents dans cette région (Figure I.7-A). Les autres espèces bien modélisées sont les arbres dominants des forêts catalanes, se développant dans des habitats relativement extrêmes où l'impact humain est traditionnellement faible. C'est le cas des forêts sub-alpines de *Pinus uncinata* et d'*Abies alba* ou à un moindre degré des forêts de *Fagus sylvatica* en montagne, toutes situées dans les zones froides et humides des faces nord, inappropriées à l'agriculture. Cette conclusion est tout aussi valable pour *Quercus suber* et *Castanea sativa* dans la région d'étude, qui se développent sur des sols pauvres et siliceux, qui sont inadéquats aux cultures.

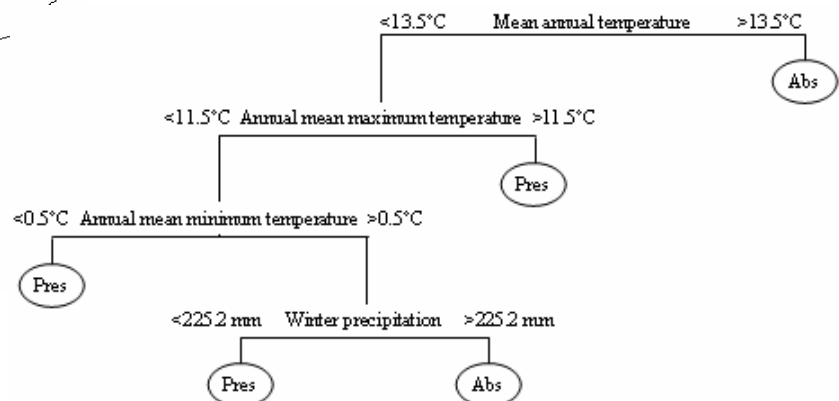
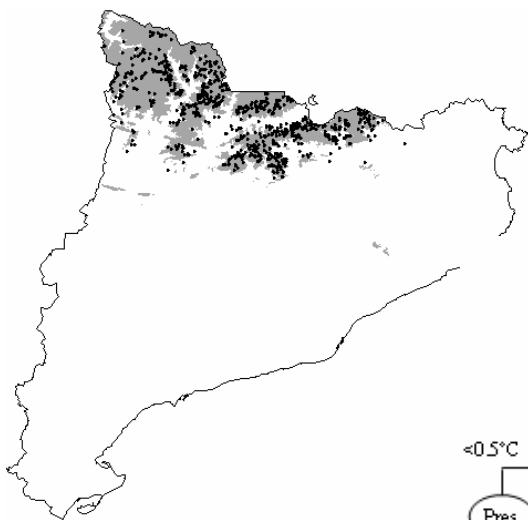
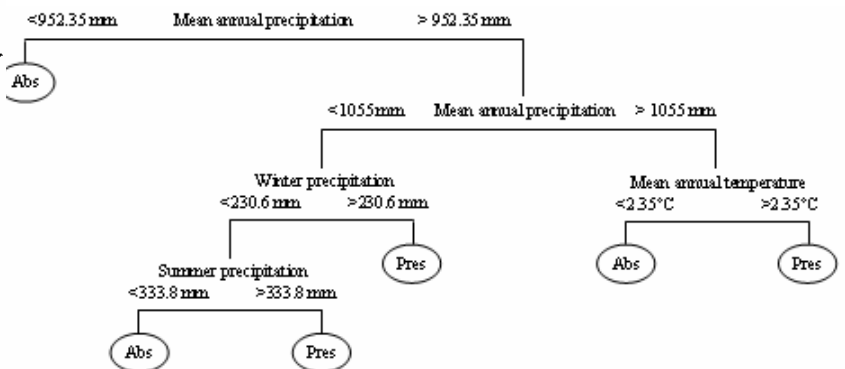
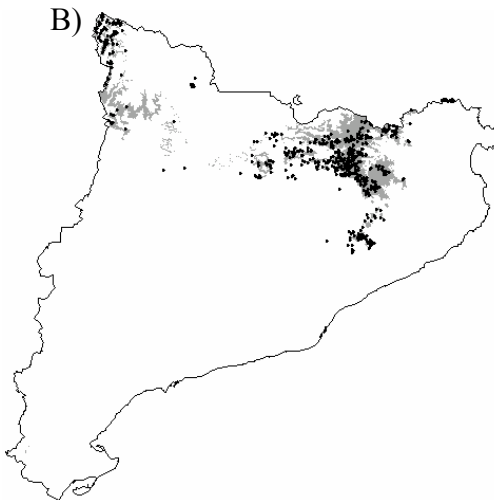
La fiabilité des modèles est plus faible pour les espèces restantes. Parmi elles, les forêts de pins (*Pinus nigra*, *P. pinaster*, *P. pinea* et *P. sylvestris*) sont modélisées avec une relativement bonne précision, indiquant qu'elles ont été plantées ou se sont dispersés principalement à l'intérieur de leur zones potentielles de croissance. Bien que nous aurions supposé que leurs distributions actuelles seraient fortement influencées par l'histoire de l'utilisation des terres et par la gestion actuelle, en l'absence de ces données les modèles bioclimatiques-géologiques semblent suffisants. Deux raisons peuvent expliquer ce résultat : premièrement, ces espèces commerciales ont eu l'opportunité de se disperser dans tous les sites adéquats à travers leur plantation. Par conséquent, le climat et la géologie peuvent procurer un substitut raisonnable des effets de la gestion forestière. Deuxièmement, les espèces ont été plantées dans des régions connues pour être climatiquement et géologiquement adéquates. Cependant, les modèles détectent des occurrences (fausses négatives) de *Pinus sylvestris* et *P. nigra* dans le nord-est de la Catalogne (climat maritime), où les espèces ont été plantées en dehors de leurs distributions potentielles. Ces fausses négatives reflètent typiquement l'incapacité des modèles statiques tels que CTA, de prendre en compte les facteurs historiques et des situations dynamiques impliquées dans le cas d'espèces plantées en dehors de leurs niches écologiques (Guisan & Zimmermann 2000).

Figure I.7 : Distributions observées (points noirs), potentielles (gris) et arbres de classification correspondants du A) pin d'Alep (*Pinus halepensis*), B) hêtre (*Fagus sylvatica*) et pin à crochets (*Pinus uncinata*)

A)



B)



Les espèces sous-dominantes eurosibériennes qui répondent aux facteurs extrazonaux ou azonaux, sont plus ou moins bien modélisées et prédites, indiquant une dépendance variable aux conditions de milieu. Quelques unes de ces espèces (*Populus tremula*, *Sorbus aria* et *Ilex aquifolium*) se trouvent de préférence dans de vastes forêts au sud de la Catalogne (Pino *et al.* 2000), en marge de leur aire de répartition européenne, où seulement des populations localisées persistent. Dans ces zones, elles ont un comportement d'espèces de forêts intérieures (Forman & Godron 1986) où elles sont protégées des interactions avec les habitats humains adjacents. Sinon, la faible adéquation des modèles pour ces espèces azonales peut être le résultat de la stratégie d'échantillonnage de l'IEFC. En échantillonnant seulement les espèces les plus abondantes dans chaque quadrat, les espèces les plus rares, dominées par les arbres de fin de succession, ont probablement été sous-échantillonnées. Cette explication est aussi plausible pour des espèces avec des distributions clairsemées en Catalogne comme *Sorbus aria*.

Pour finir, les espèces comme *Quercus ilex* et *Q. humilis*, qui ne sont pas très bien modélisées, devraient être respectivement la végétation dominante dans les zones non siliceuses des plaines de Catalogne et sud de la France, et la végétation dominante des zones sub-méditerranéennes bordant la région eurosibérienne (Bolòs *et al.* 1990). Cependant la majorité de leurs habitats potentiels ont été substitués par des forêts de pins, arbustes, prairies, cultures et zones urbaines. Ce fait est certainement la principale raison de la faible adéquation des modèles pour ces espèces.

3.5 Choix des variables explicatives

Basé sur Annexe 1.5 et 1.6 :

Thuiller, W., Araújo, M. B. & Lavorel, S. (2004). Do we need land-cover data to predict species distributions? *Journal of Biogeography*, sous presse

Araújo, M. B., Thuiller, W., Williams, P. H., & Reginster, I. (2004). Downscaling European species atlas distributions to a finer resolution: implications for conservation planning, *Glob Ecol & Biog*, sous presse

Le choix des variables explicatives ou prédictives à un rôle primordial dans la compréhension, l'estimation, et la prédiction des distributions d'espèces. D'un point de vue mécaniste, il est recommandé de prédire la distribution spatiale d'entités biotiques sur la base de paramètres écologiques qui sont supposés être les facteurs limitants de la réalisation de fonctions physiologiques et de survie à long terme. Les modèles basés sur les niches tentent de relier la distribution actuelle observée d'entités biologiques avec les paramètres abiotiques de la zone étudiée. Ceci suppose donc de disposer de variables suffisamment pertinentes pour

rendre compte des corrélations entre le milieu et la distribution. A échelle biogéographique ou continentale, il est reconnu que la répartition des plantes est principalement dépendante des conditions climatiques (Woodward 1987; Stephenson 1990), et principalement de la température et des précipitations (Pigott & Pigott 1993). Bien que cela reste beaucoup moins documenté, il est évident qu'il en va de même pour de nombreux animaux (Budyko 1974; Brown 1995). Cependant, plus l'échelle diminue, plus l'importance d'autres facteurs physiques peut augmenter (Willis & Whittaker 2002). L'utilisation des terres par l'homme est reconnue comme un facteur responsable de la restriction, de la migration ou tout simplement de la disparition de nombreuses espèces (Vitousek *et al.* 1997; Mooney & Hobbs 2000). Il est facilement concevable qu'il n'y ait pas une frontière fixe entre l'importance du climat à grande échelle et l'importance de l'utilisation du sol ou de la fragmentation à échelle plus fine. Dans les zones de forts gradients climatiques (ex : Corse ou Catalogne), le climat joue un rôle très fort sur la distribution des espèces même à échelle très locale et peut être largement prépondérant sur des variables de végétation ou de structuration du paysage.

3.5.1 Variables directes *versus* variables indirectes

Austin & Smith (1989) définissent trois types de variables de milieu jouant sur la distribution des espèces : variables directes, indirectes et ressources. Les variables directes sont celles qui ont un impact sur la physiologie des espèces et leur succès reproducteur (ex : température, degré-jour de croissance, indice d'humidité, précipitation). Les variables indirectes sont celles qui n'ont pas d'impact direct sur la performance des espèces, mais qui expliquent simultanément différentes variables de ressources et variables directes (ex : pente, altitude, orientation, sols, type d'habitat). Enfin les variables dites de ressources représentent ce qui est consommé ou utilisé par les espèces (ex : nourriture, lumière).

En modélisation statique, ce sont les variables directes et indirectes qui sont utilisées (Guisan & Zimmermann 2000). Le biais principal apporté par les variables indirectes (ou dites de substitution) est qu'un modèle les utilisant peut uniquement être appliqué à l'intérieur de l'étendue géographique et pour la période sur laquelle il a été calibré sachant que les jeux de corrélation entre variables pourraient changer (Guisan & Zimmermann 2000). En effet, dans une région différente, la même variable indirecte pourra refléter une combinaison de variables directes autres que celles pour laquelle elle a été utilisée. C'est typiquement le cas des variables topographiques comme l'altitude qui est un substitut de variables directes comme la température et la précipitation. Cependant, la combinaison de ces variables dans les Alpes est très probablement différentes de celles dans les Pyrénées ou dans les Balkans et a peu de

chance de se maintenir dans un contexte de changement climatique. Inversement, les variables directes sont les plus aptes à être utilisées quelle que soit la région étudiée, car selon la théorie de la niche, les espèces ont un optimum et une tolérance ‘absolue’ en fonction de ces facteurs, indépendamment de la localisation spatiale et de la topographie de la zone. Par exemple, dans le reste de cette thèse, j’ai utilisé des données climatiques fournies dans le cadre du projet européen ATEAM comme la température du mois le plus froid, les degré-jours de croissance ou un index d’humidité. Pour les plantes, la température du mois le plus froid doit discriminer les espèces sur leur capacité à assimiler et économiser l’eau du sol et les nutriments et à continuer la division cellulaire, la différenciation et croissance des tissus pour des températures très faibles. Les degré-jours discriminent les espèces sur la nécessité d’une température minimum pour compléter leur cycle de croissance (limite inférieure) et leur capacité à tolérer des excès de températures dans les tissus (limite supérieure). Enfin, l’index d’humidité discrimine les espèces selon leur perte d’eau par transpiration (limite supérieure) et leur efficacité du transport de l’eau. Les modèles qui utilisent des variables de ce type, importantes pour la physiologie et le succès reproducteur des espèces concernées, devraient être plus facilement généralisables et pourraient être utilisées pour projeter la distribution des espèces étudiées dans d’autres régions ou dans le futur. Il est d’ailleurs intéressant de noter que ce sont ces variables qui sont aussi utilisées dans la plupart des modèles mécanistes (Sitch *et al.* 2003) et non les variables indirectes.

3.5.2 Impact de l’utilisation des terres en Europe sur la distribution des espèces

L’utilisation des terres et la couverture du sol (UT/CS) est un paramètre de plus en plus important en Europe, compte tenu de l’augmentation croissante de la main mise de l’homme sur les écosystèmes naturels. L’UT/CS en elle-même et les changements d’UT/CS sont des causes importantes d’extinctions d’espèces car elles perturbent ou détruisent les écosystèmes naturels (Mooney & Hobbs 2000). Dans le monde, depuis le début du 19^{ème} siècle, l’intensification de l’agriculture, la destruction massive de forêts naturelles, la gestion des forêts ou la déprise agricole font que l’UT/CS est vue comme un facteur à part entière pour expliquer la distribution des espèces de l’échelle locale à globale (Mooney & Hobbs 2000). Si de nombreuses études ont montré l’importance de tels facteurs à l’échelle du paysage sur la distribution des espèces ou la biodiversité (Robinson & Sutherland 2002; Benton *et al.* 2003), cela n’a jamais été explicitement démontré à l’échelle biogéographique, continentale ou globale. La faible disponibilité de données explique en partie cette lacune dans les analyses. Pourtant récemment une nouvelle base de données européenne PELCOM qui cartographie

Encadré I.4

Données bioclimatiques :

Les données bioclimatiques utilisées sont des variables fournies par le Climatic Research Unit (<http://www.cru.uea.ac.uk/>) (Mitchell 2002) (<http://www.pik-potsdam.de/ateam/>): somme des précipitations annuelles, été et hiver, moyenne des températures (T) annuelles, minimum des T du mois le plus froid, degré jour de croissance > à 5° (Prentice et al 1992) et index d'humidité (ratio évapotranspiration réelle sur potentielle).

Données utilisation des terres :

Les données UT sont utilisées à une résolution de 10' basée sur la fenêtre géographique du projet ATEAM. Les données sous forme de grille ont été agrégées de la base de données PELCOM. Cette base a été développée principalement à partir de données de télédétection. PELCOM a été sélectionné à cause de sa couverture complète de l'Europe et de l'homogénéité de la méthodologie pour la classification des utilisations des terres (Voir Annexe x et site pour plus de détail). Les données comprennent quatre grandes classes d'UT : forêts, agriculture, urbanisation (perurban) et autres. La classe Forêts a été subdivisée en trois sous-classes : % de décidus (perdeci), % de conifères (perconi) et % de forêts mixtes (permixf). La classe Agriculture a elle aussi été subdivisée en trois classes : % de terres arables (perara), % de prairie (pergrass) et % de cultures permanentes (perpermc). La classe Autre comprend le % d'arbustes (pershrub), le % de toundra (perbarrenl), le % d'eau intérieure (perinwater), le % de terres marécageuses (perwetland), le % de glace et neige permanente (perpice) et le % de mer (persea). Ces pourcentages correspondent au % de chaque classe dans chaque cellule. Comme pour la suite, nous utilisons des données espèces à 50x50 km, les données 10° ont été agrégées par simple moyenne.

Méthode :

Analyse de co-inertie 'Outlying Mean Index' (OMI) récemment proposée par Doledec *et al.* 2000) et complétée par Dray *et al.* In press). Cette analyse est similaire à l'analyse canonique des correspondances (ACC : ter Braak 1986), mais permet de séparer les classes d'utilisation des terres et de mesurer la distance entre les conditions moyennes bioclimatiques utilisées par chaque classe d'UT/CS et la moyenne des conditions bioclimatiques à travers l'Europe. De plus, OMI à l'avantage de ne faire aucune hypothèse quant à la forme des courbes de réponses des classes d'UT/CS aux variables bioclimatiques (Dray *et al.* In press).

différentes classes d'UT/CS a été créée (Mücher 2000) et mise à disposition sur le site Internet : <http://cgi.girss.wageningen-ur.nl/cgi/projects/eu/pelcom/index.htm>.

3.5.2.1 Déterminants climatiques de l'utilisation des terres

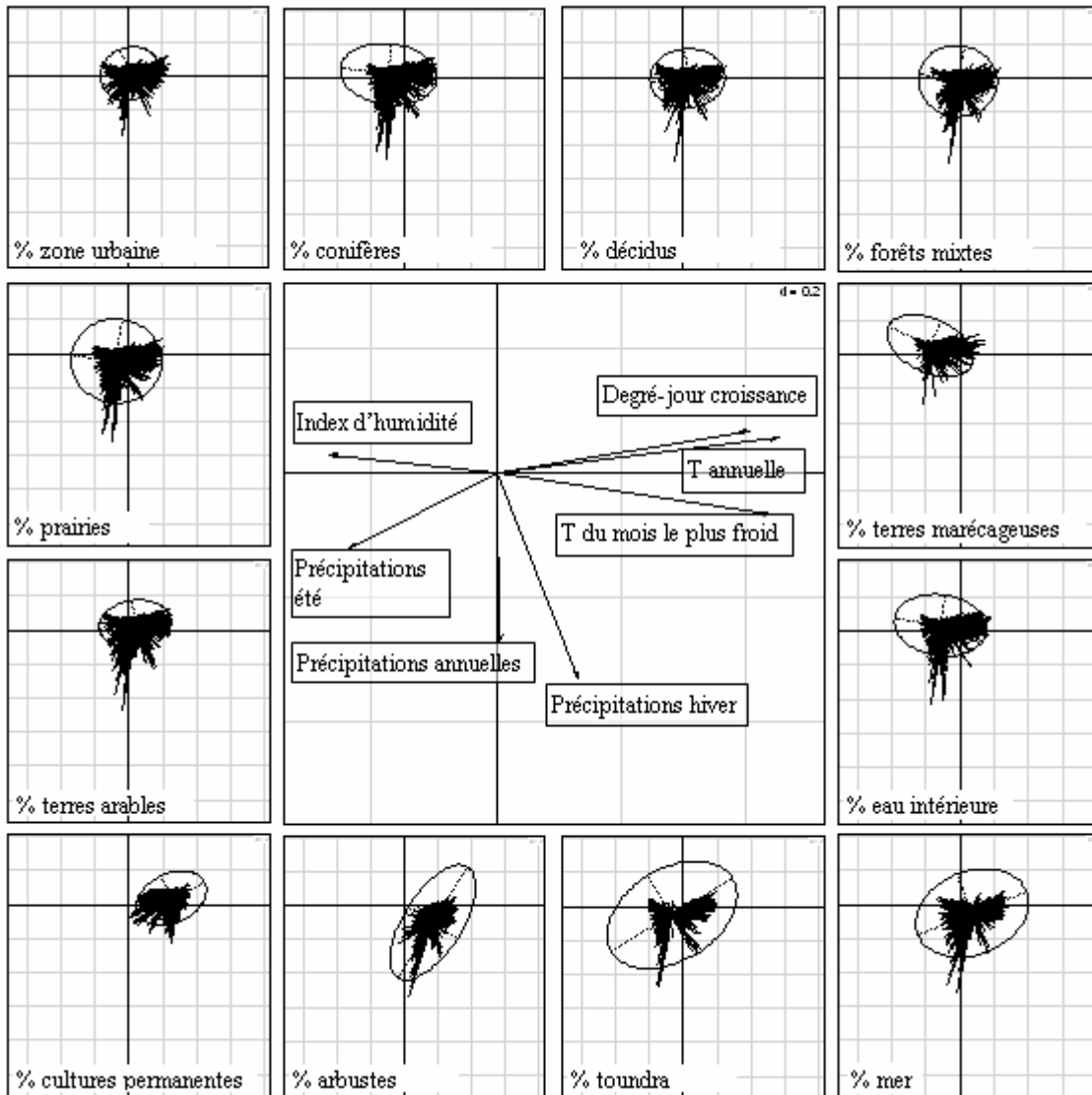
Comme le climat contrôle de façon globale la distribution des êtres sur terre, il est aussi probable qu'il contrôle plus ou moins directement la distribution de l'UT/CS à échelle européenne. Pour vérifier cette hypothèse, j'ai étudié dans quelle proportion l'UT/CS en Europe pouvait être expliquée par le climat (Encadré I.4).

L'analyse de co-inertie OMI (Doledec *et al.* 2000) permet de déterminer les 'niches' climatiques de la couverture du sol et utilisation des terres. L'espace climatique créé par OMI est constitué de deux axes expliquant plus de 96% de l'inertie totale (Figure I.8). Le premier axe est fortement relié à la température (65%), tandis que le second explique un gradient de précipitation (31%). La projection des variables UT et CS sur cet espace climatique à deux dimensions montrent que bien qu'un certain nombre de variables soient d'origine anthropique (ex : culture permanente), le climat permet d'expliquer leur distribution à grande échelle. Par exemple, on peut noter que les cultures permanentes se situent principalement dans des régions tempérées, alors que les prairies se retrouvent principalement dans des zones de moyenne montagne. A l'inverse, certains types de CS se retrouvent dans toutes les zones de l'espace bioclimatique européen. C'est le cas par exemple de la mer, des eaux intérieures (rivières, lacs), des terres arables, des zones urbaines ou des forêts de conifères.

Ceci est un résultat très important en termes de modélisation de la distribution des espèces à grande échelle. En effet, ce résultat explique pourquoi inclure les deux types d'information (UT/CS et climat) dans un même modèle de distribution d'espèces, peut amener une redondance de l'information. Le degré par lequel ce résultat est une conséquence de l'échelle spatiale et de la résolution grossière reste inconnu. Toutefois, il est possible que les modèles utilisant des données plus localisées avec une résolution fine produisent des patrons différents (Collingham *et al.* 2000). En effet, un schéma hiérarchique de contrôle du milieu sur la distribution des espèces a été suggéré, dans lequel les variables climatiques étaient des déterminants à grande échelle, suivis par la géologie, la couverture végétale, l'utilisation des terres et la topographie, qui modèrent les effets des variables 'macro' climatiques (Franklin 1995).

Dans notre analyse, bien qu'une importante proportion de la variabilité dans la distribution de l'UT/CS soit expliquée par le climat, il reste des résidus se rapportant à des

Figure 1.8 : Corrélation canonique des variables bioclimatiques sur le plan de l'OMI. Représentation 2D de la 'niche' des variables d'utilisation des terres sur ce même plan. Une ellipse regroupe 90% des points. Le centroïde de l'ellipse représente la position moyenne d'une classe d'utilisation des terres dans l'espace bioclimatique.



patrons plus régionaux non expliqués par le climat. Quelques variables d'UT/CS sont peu ou pas expliquées par les facteurs climatiques, tels que les eaux intérieures, la mer, les terres arables ou les forêts de conifères. La mer est présente autour d'une grande proportion du continent Européen et est évidemment indépendante du climat. Plus surprenant, la distribution des forêts de conifères ne paraît pas être influencée par le climat. Une raison plausible pour expliquer ce phénomène est le fait que cette classe englobe une grande variété d'espèces aux exigences climatiques et écologiques très différentes, ce qui peut expliquer qu'il est difficile de trouver un patron général (Richardson 1998). Par exemple, le genre *Pinus* est présent dans les plaines sibériennes (*Pinus sibirica*), dans les Alpes (*Pinus cembra*), dans des zones restreintes de la région méditerranéenne (*Pinus halepensis*) ou dans les régions euro-sibériennes (*Pinus sylvestris*). Une telle variété de chorotypes explique pourquoi la distribution des forêts de conifères ne semble pas être contrôlée par le climat dans notre analyse, bien que les espèces individuelles de cette classe soient fortement liées aux conditions climatiques (Richardson 1998; Thuiller *et al.* 2003). Ces classes d'UT/CS peu ou pas reliées au climat à l'échelle européenne, sont cependant influencées par de multiples autres facteurs régionaux ou locaux tels que le type de sols, la pente, l'exposition ou la disponibilité des eaux souterraines, qui ne sont pas pris en compte dans notre analyse. Le sol, par exemple, est un déterminant crucial de la distribution de certaines espèces (ex : *Quercus suber*, Thuiller *et al.* 2003), et devrait probablement expliquer une part notable des résidus laissés par le climat dans cette analyse.

3.5.2.2 Importance de l'UT/CS pour prédire la distribution des espèces

Etant donné ce patron de co-variation entre les variables UT/CS et les variables bioclimatiques, il est évident qu'inclure directement les variables bioclimatiques et les variables UT/CS dans un modèle de distribution d'espèces avec sélection de variables, revient à privilégier les variables bioclimatiques. Cependant, les variables d'UT/CS peuvent quand même contribuer significativement aux modèles de distribution d'espèces en Europe, car des patrons régionaux d'UT/CS, correspondants à la part résiduelle de l'OMI, peuvent influencer les distributions d'espèces particulières. Pour tester quelle peut être l'influence de l'UT/CS pour prédire la distribution des espèces à grande échelle tout en tenant compte de l'importance prépondérante du climat, nous avons mis en place une méthode itérative pour prendre en compte ces deux composants de manière hiérarchique. Pour que notre analyse aient une valeur générale, nous avons modélisé la distribution de 2196 espèces européennes, correspondant à

Encadré I.5

Données bioclimatiques : les mêmes que dans l'Encadré I.4

Données utilisation des terres/couverture du sol : les mêmes que dans l'Encadré I.4

Données espèces : Dans cette analyse, nous avons utilisé les quatre grands atlas européen cartographiant respectivement la distribution des plantes (Atlas Florae Europaea : Jalas & Suominen 1972-1996), des mammifères (Mitchell-Jones *et al.* 1999), des oiseaux nicheurs (Hagemeijer & Blair 1997) et des reptiles et amphibiens (Gasc *et al.* 1997). Après avoir enlevé les espèces avec un faible nombre d'occurrences (<10), 2196 espèces ont été retenues.

Modèles : La distribution des 2196 espèces a été modélisée en utilisant des GAM d'une manière itérative (Tableau I.6). Cette approche permet de mesurer la puissance explicative (en terme de déviance expliquée) et la puissance prédictive (en terme d'augmentation d'AUC) des variables d'UT/CS. Les modèles ont été calibrés sur un échantillonnage aléatoire (70% des données originales) et ensuite évalué à l'aide la courbe ROC sur les données d'évaluation.

Ensuite, le meilleur modèle (bioclimatique ou alternatif) pour chaque espèce a été utilisé pour projeter la distribution des espèces à une résolution plus fine de 10'.

Tableau I.6. Description du processus itératif.

Pas	Processus itératif
1	GAM avec procédure de sélection de variables en utilisant seulement les variables bioclimatiques (GAM-clim)
2	GAM avec procédure de sélection de variables sur les résidus du pas précédent en utilisant uniquement les variables d'utilisation des terres
3	GAM en utilisant les variables bioclimatiques (pas 1) et utilisation des terres (pas 2) (GAM-clim-UT)
4	Evaluation de la puissance prédictive de GAM-clim en utilisant la courbe ROC
5	ANOVA comparative entre le GAM-clim et le GAM-clim-lu (Test)
6	- Si Test significatif : évaluation de GAM-clim-lu sur données d'évaluation avec courbe ROC - sinon, évaluation de GAM-clim sur données d'évaluation (même que Pas 3)
7	Moyenne courbe ROC des GAM-clim sur toutes les espèces
8	Moyenne courbe ROC des GAM-clim-lu (si Test significatif) ou GAM-clim (si Test non significatif)
9	Modèle GAM-clim-lu ou GAM-clim utilisé pour projeter les distributions des espèces à 10° selon leur courbe ROC

165 mammifères, 117 reptiles et amphibiens, 387 oiseaux nicheurs et 1527 plantes (Encadré I.5, Tableau I.6).

Outre le besoin d'évaluer la puissance prédictive et explicative amenées par les variables l'UT/CS aux modèles bioclimatiques, cette analyse permet de simuler la distribution de toutes ces espèces à 10', une échelle plus fine que les atlas originellement à 50x50 km. En effet, une des limites des atlas de distribution pour les plans de conservation est leur résolution grossière. Ce processus itératif pourrait permettre de produire les cartes de distribution des atlas à une échelle plus fine et d'augmenter l'applicabilité des plans de conservations (Barbosa *et al.* 2003).

Ce processus itératif a été conçu tout d'abord pour modéliser la distribution de chaque espèce en utilisant uniquement les variables bioclimatiques (modèles bioclimatiques), et ensuite pour tester si l'apport des variables UT/CS diminuait la part résiduelle des modèles bioclimatiques. Si oui, les variables UT/CS sélectionnées sont incluent dans les modèles bioclimatiques créant ainsi des modèles dits alternatifs. La puissance explicative des variables UT/CS est mesurée par la part supplémentaire de déviance expliquée apportée aux modèles bioclimatiques, et la puissance prédictive des variables UT/CS est évaluée en comparant celle des modèles bioclimatiques à celle des modèles alternatifs.

Puissance explicative. L'inclusion de l'UT/CS a, en moyenne, augmenté la puissance explicative pour tous les groupes (Tableau I.7). Ceci est particulièrement notable pour les mammifères, les oiseaux et les arbres, pour lesquels l'inclusion des variables UT/CS augmente la déviance expliquée des modèles bioclimatiques dans plus de 60% des espèces. Dans le but d'explorer les effets additionnels de l'UT/CS après que les modèles bioclimatiques aient été ajustés, nous avons sélectionné quelques espèces ayant des relations démontrées avec des variables d'UT/CS à échelle plus fine.

C'est le cas du loup (*Canis lupus*), pour lequel sa distribution géographique est inversement corrélée avec la population et les activités humaines (Breitenmoser 1998; Massolo & Meriggi 1998). Dans notre analyse, les variables d'UT/CS retenues pour cette espèce sont les pourcentages de zones urbaines et de prairies. Les courbes de réponse de ces deux variables sont respectivement négative exponentielle et négative linéaire, montrant que malgré la résolution grossière et l'échelle européenne, les variables d'UT/CS importantes pour le loup peuvent être détectées par les modèles. Toutes espèces de mammifères confondues, les trois variables les plus retenues par les modèles sont les pourcentages de zones arables, de cultures permanentes et de forêts mixtes. Robinson, Wilson *et al.* (2001) ont montré que la perdrix

Tableau I.7 : Puissance explicative des variables d'utilisation des terres et de couverture du sol. *n* est le nombre d'espèces. Pas 2 correspond au nombre d'espèces pour lesquelles les variables UT/CS ont été retenues comme diminuant la part résiduelle des modèles bioclimatiques (Pas 2, Tableau I.6). Pas 5 correspond au nombre d'espèces pour lesquelles les modèles alternatifs ont une plus haute déviance expliquée que les modèles bioclimatiques (Pas 5, Tableau I.6). Pourcentage correspond au pourcentage de cas où l'inclusion des variables UT/CS augmente la puissance prédictive des modèles bioclimatiques.

	<i>n</i>	Pas 2	Pas 5	Pourcentage
Mammifères	165	162	107	65%
Oiseaux	387	377	241	62%
Arbres	71	71	43	61%
Arbustes	78	78	43	55%
Reptiles/Amphibiens	117	110	62	53%
Herbes	1378	1265	624	45%

(*Perdix perdix*) avait décliné en Angleterre ces dernières années, conséquence de l'intensification des pratiques pastorales et de la baisse des terres arables disponibles. De même, dans notre analyse, la variable qui augmente la déviance expliquée du modèle bioclimatique de cette espèce est le pourcentage de terres arables en Europe. En ce qui concerne tous les oiseaux, l'agriculture (% de zones arables et de cultures permanentes) et la mer sont les facteurs qui réduisent principalement la part résiduelle des modèles bioclimatiques. Pour les arbres, l'exemple choisi est celui du chêne sessile (*Quercus petraea*), un chorotype euro-sibérien ayant une vaste distribution (Tutin *et al.* 1964-1993). La seule variable d'UT/CS sélectionnée pour cette espèce, est le pourcentage de décidus, avec une forte relation positive. Cette espèce est en effet un décidu dominant dans les forêts d'Europe centrale. En général, pour les arbres, il n'y a pas de variables significativement plus sélectionnées que d'autres, bien que le pourcentage de mer, d'eaux intérieures, et d'arbustes apparaissent comme des variables prépondérantes. Pour les reptiles et amphibiens, l'inclusion des variables d'UT/CS augmente la puissance explicative des modèles bioclimatiques dans plus de la moitié des cas (Tableau I.7). Ces espèces répondent plus particulièrement aux pourcentages de décidus, de conifères et de terres arables. Ces observations sont conformes avec des observations publiées. Par exemple, Martín & Lopez (2002) ont montré que le lézard (*Lacerta lepida*) se localisait principalement sous des forêts de chênes et sous-bois arbustifs, mais évitait les cultures permanentes. Ici, l'analyse a retenu le pourcentage de forêts mixtes (relation positive) et de cultures permanentes (relation négative).

Comme certaines classes d'UT/CS largement répandues ne sont pas restreintes par le climat, il n'est donc pas surprenant qu'elles soient fréquemment retenues par les modèles GAM (Tableau I.6 - Pas 2), et qu'elles diminuent significativement la part résiduelle des modèles bioclimatiques pour beaucoup d'espèces. Les terres arables et cultures permanentes augmentent significativement la puissance explicative des modèles bioclimatiques pour de nombreux oiseaux et mammifères. Nos résultats corroborent donc des études montrant l'importance positive ou négative des terres arables pour les oiseaux, mammifères et plantes, de l'échelle régionale à locale (ex. Robinson *et al.* 2001; Benton *et al.* 2003). Ces auteurs montrent que l'intensification de l'agriculture et la contraction des cultures arables ont produit des extinctions locales très importantes chez les oiseaux et mammifères. Notre étude indique donc que même à grande échelle et résolution grossière, les variables induites par l'homme peuvent dans certains cas être utiles pour décrire et expliquer la distribution des espèces (ex : *Perdix perdix*).

Tableau I.8 : Puissance prédictive des modèles bioclimatiques et alternatifs mesurée par AUC (Encadré I.1). EcT, Min, Moy et Max sont respectivement le minimum, la moyenne, le maximum AUC à travers toutes les espèces. Diff correspond à la différence entre la moyenne AUC des deux approches. Cali et Eval correspondent respectivement aux données de calibration et d'évaluation.

	Modèles bioclimatiques				Modèles alternatifs				Diff	
	EcT	Min	Mean	Max	EcT	Min	Mean	Max		
Reptiles	Cali	0.0467	0.8017	0.9573	0.9999	0.0413	0.8099	0.9627	0.9999	-0.0054
Amphibiens	Eval	0.0505	0.7791	0.9497	0.9988	0.0487	0.7933	0.9504	0.9988	-0.0007
Mammifères	Cali	0.0582	0.7457	0.9277	0.9992	0.0522	0.7812	0.9358	0.9992	-0.0081
	Eval	0.0632	0.7267	0.9140	0.9987	0.0603	0.7552	0.9133	0.9975	0.0007
Oiseaux	Cali	0.0459	0.7853	0.9388	0.9990	0.0421	0.8227	0.9447	0.9990	-0.0059
	Eval	0.0525	0.7592	0.9254	0.9980	0.0512	0.7592	0.9260	0.9980	-0.0006
Herbes	Cali	0.0419	0.7341	0.9568	1.0000	0.0419	0.7341	0.9567	1.0000	0.0001
	Eval	0.0503	0.6628	0.9313	1.0000	0.0503	0.6628	0.9313	1.0000	0.0000
Arbustes	Cali	0.0314	0.8692	0.9708	1.0000	0.0256	0.8950	0.9743	1.0000	-0.0035
	Eval	0.0421	0.8041	0.9530	1.0000	0.0406	0.8046	0.9558	1.0000	-0.0028
Arbres	Cali	0.0277	0.8950	0.9642	0.9993	0.0327	0.8482	0.9615	0.9993	0.0027
	Eval	0.0347	0.8188	0.9436	1.0000	0.0384	0.8136	0.9414	1.0000	0.0032

Cependant, notre approche est uniquement corrélative et la signification des relations observées entre les variables UT/CS et les distributions des espèces n'est pas toujours évidente. Bien que les modèles bioclimatiques identifient aussi des relations corrélatives, il existe de nombreuses études qui ont examiné les relations causales entre la survie des espèces et des variables bioclimatiques (au moins pour les plantes) (Woodward 1987, 1992). Ce genre d'analyses n'a pas été développé à grande échelle pour les relations entre UT/CS et la survie des espèces. La relation causale entre les variables d'UT/CS et la distribution des espèces n'est pas simple à découvrir car souvent indirecte. C'est le cas par exemple pour les plantes, où l'utilisation des terres ne peut être considérée comme un facteur ayant un impact physiologique direct. Cependant, pour les vertébrés, la relation peut être plus directe. Par exemple, les habitats de bonne qualité pour les oiseaux de cultures « farmland birds (stubble, game cover, game feeders) » se situent principalement dans des zones de terres arables (Benton *et al.* 2003). Les passereaux granivores et petits rongeurs sont évidemment fortement liés aux cultures permanentes et terres arables où ils trouvent refuges et nourriture.

Puissance prédictive. La puissance prédictive des variables UT/CS a été évaluée en analysant les courbes ROC sur les données d'évaluation (Tableau I.8). En moyenne, les modèles bioclimatiques et alternatifs ont un fort pouvoir prédictif. Reptiles/amphibiens et arbustes sont les espèces les mieux prédites avec en moyenne une AUC supérieure ou égale à 0,95. Les mammifères ont la plus faible puissance prédictive (bien que toujours haute). Les prédictions en utilisant les modèles bioclimatiques ou les modèles alternatifs montrent des résultats très similaires. Parmi les exemples présentés précédemment, il n'y a aucune espèce pour laquelle l'inclusion des variables d'UT/CS augmente significativement la puissance prédictive des modèles bioclimatiques. Néanmoins, l'inclusion de ces variables augmente la puissance prédictive des modèles pour 34% des mammifères, 40% des reptiles, 34% des oiseaux et 21% des plantes. Les animaux apparaissent être le plus corrélés aux variables d'UT/CS, mais les différences de puissance prédictive entre les modèles bioclimatiques et d'UT/CS sont trop faibles pour tirer des conclusions claires sur le bénéfice d'inclure les variables d'UT/CS pour prédire la distribution des animaux à échelle continentale.

Quelques hypothèses peuvent être avancées pour expliquer ce manque d'efficacité des variables UT/CS :

- La mesure d'adéquation (« Goodness of fit ») ne permet pas d'évaluer la puissance prédictive des modèles (Hosmer & Lemeshow 1989). Même si quelques variables d'UT/CS

augmentent la puissance explicative des modèles bioclimatiques, cela n'implique pas qu'elles augmentent aussi la puissance prédictive.

- Le climat explique presque toute la variation spatiale de l'UT/CS et les résidus ne sont pas assez importants pour augmenter la puissance prédictive des modèles. Les résidus des relations climat-UT/CS peuvent être fortement liés à des variations régionales et ne représentent pas de gradient affectant la distribution des espèces suffisamment fortement.
- Les variables d'UT/CS sont indirectes et sont supposées avoir une faible pertinence physiologique pour les espèces, donc elles devraient seulement être appliquées dans une étendue géographique restreinte, parce que dans une région ou période de temps différentes, la même variable d'UT/CS peut correspondre à différentes combinaisons de gradients directs et de ressources (Guisan & Zimmermann 2000).

En utilisant une grille 50x50 km à travers l'Europe, nous avons démontré que l'UT/CS est principalement dirigé par le climat. Cependant, à l'intérieur de chaque pixel, il y a une forte hétérogénéité sous-jacente qui est perdue (voir la base PELCOM), alors que cette hétérogénéité est un facteur crucial pour prédire l'occurrence des espèces à l'échelle du paysage (Cowling & Lombard 2002). Cette contrainte due aux données engendre deux problèmes : premièrement, le problème d'utiliser des données avec une résolution insuffisante pour expliquer les patrons observés (Robinson *et al.* 1992; Pearson 1993), deuxièmement, la possibilité que les variables d'UT/CS utilisées dans cette étude n'incluent pas toutes les variables pertinentes nécessaires. Par exemple, les variables représentant des indices d'intensification agricole (Mader 1990), d'hétérogénéité d'habitat (Bascompte & Solé 1996; Allen & O'Connor 2000), de fréquence et d'intensité des perturbations (Croonquist & Brooks 1993) ou de densité de routes (Mader 1990) pourraient être plus pertinentes pour augmenter la capacité prédictive des modèles purement bioclimatiques, même à une échelle de résolution 50x50 km.

4 Conclusions

Pour conclure cette première partie, on peut simplement rappeler quelques résultats importants :

1 - Il existe différents types de modèles utiles pour expliquer et prédire la distribution d'organismes. Les modèles mécanistes, basés sur les processus, permettent de mesurer et prédire la niche fondamentale des organismes sélectionnés (définie comme un hypervolume à n -dimensions dans lequel l'organisme peut exister indéfiniment). Ces modèles utilisent des variables de milieu dites de ressources et de réponses physiologiques à la température, et nécessitent donc de grandes quantités de données, malheureusement manquantes pour un grand nombre d'espèces. L'alternative à ces modèles sont les modèles basés sur la représentation de la niche réalisée des organismes à l'aide de variables de milieu. L'hypothèse forte est la condition de quasi-équilibre des organismes étudiés avec leur milieu. Les modèles basés sur les niches permettent de traiter un grand nombre d'espèces sans en avoir une connaissance illimitée. J'ai utilisé ce type de modèle durant cette thèse.

2 - Parmi les modèles basés sur les niches, différentes approches sont disponibles, selon l'objectif principal de l'étude (explicatif vs prédictif), la qualité des données (présence-absence vs présence seulement vs abondance) et leur facilité d'interprétation (boîte noire). Les modèles additifs généralisés (GAM), les modèles linéaires généralisés (GLM) et les arbres de classifications (CTA) offrent un bon compromis entre leur pouvoir explicatif et prédictif sur des données présence-absence. Par ses fonctions non paramétriques, GAM procure un outil intéressant pour étudier les courbes de réponse des organismes aux variables de milieu, et ce à tout type d'échelle.

3 - La qualité des données d'espèces est primordiale. L'absence d'une espèce dans un pixel donné est-elle due au fait que l'habitat n'y est pas favorable ou de la non-observation ? Il existe des méthodes spécialement conçues pour traiter les données de présences seulement. Cependant, bien qu'elles soient fiables et potentiellement aussi prédictives qu'un GLM, ces méthodes ont tendance à surestimer grandement la distribution actuelle des organismes étudiés. Bien que cette surestimation, dans le cas de mise en place d'un plan de conservation spécifique ou dans le cas de l'étude de prolifération d'espèces invasives, soit potentiellement utile, il s'avère que l'absence d'une espèce (à grande échelle) apporte une information non négligeable sur la qualité des habitats et sur l'aptitude des espèces à s'y installer ou non.

4 - Les espèces en marge de leur aire de distribution sont potentiellement plus difficiles à modéliser, car elles répondent à des facteurs de milieu souvent azonaux, topographiques et locaux, permettant leur persistance sur les bornes de leur distribution. Inversement, des espèces au centre de leur aire de distribution sont plus facilement modélisées et il est aisé alors, d'étudier les déterminants écologiques responsables de leur distribution dans la zone concernée.

5 - L'écologie des organismes doit prendre une part importante dans le choix des variables de milieu et des organismes eux-mêmes. Les variables, en particulier, ne doivent pas être choisies au hasard mais selon les effets qu'elles ont sur la physiologie, la phénologie et la survie des organismes étudiés. On a pu démontrer qu'à échelle européenne les variables macro-climatiques permettaient d'expliquer presque totalement la distribution et la structuration de l'utilisation des terres et de la couverture du sol. La prédominance forte de ces variables macro-climatiques à grande échelle et large résolution rend leur utilisation quasi-nécessaire pour la prédiction de tout type d'espèces tandis que l'utilisation des terres peut sembler utile seulement pour quelques espèces non restreintes par le climat en Europe.

2^{EME} PARTIE

PROJECTION DES FUTURS HABITATS POTENTIELS

DE LA BIODIVERSITE :

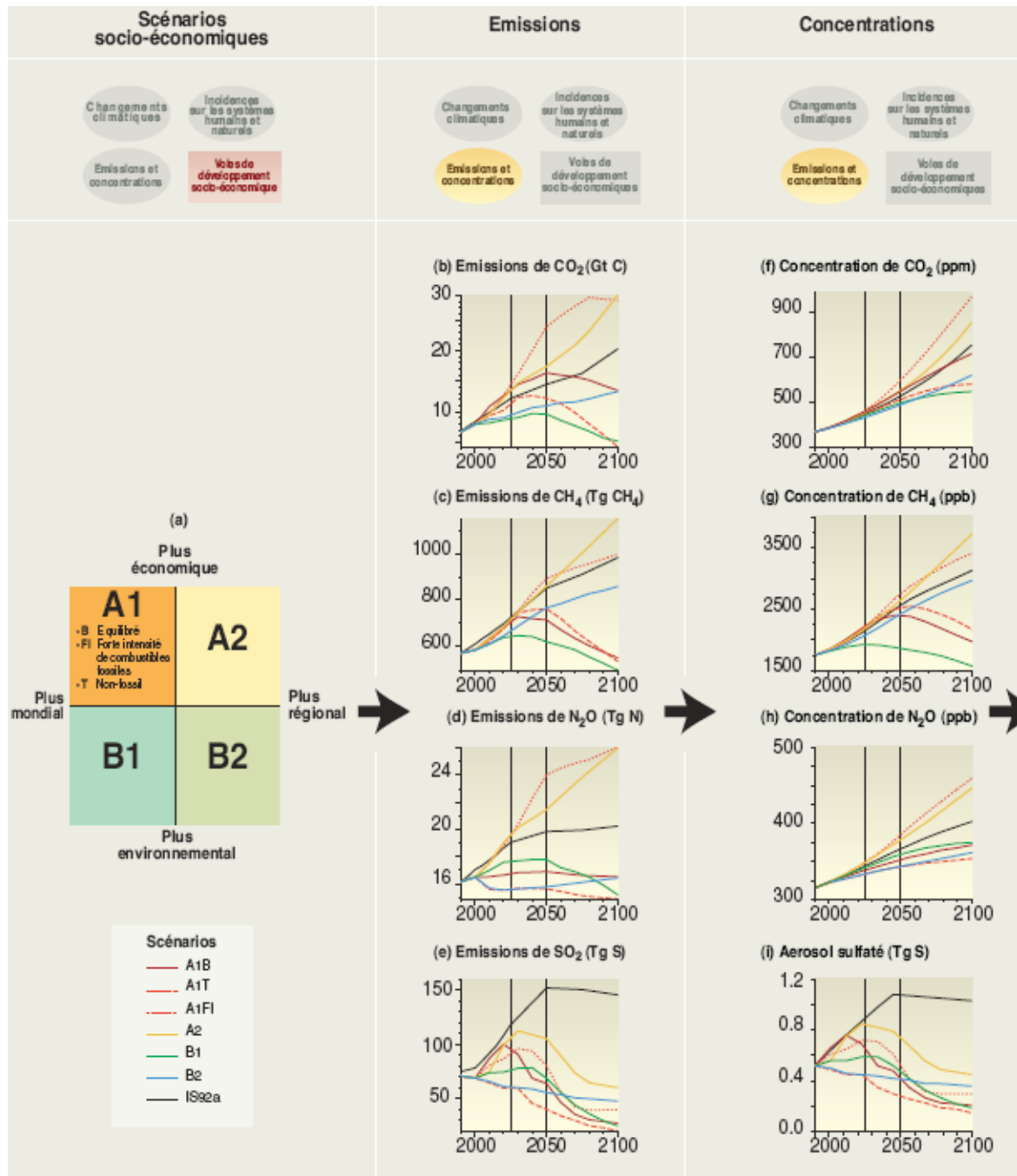
APPROCHES ET MÉTHODOLOGIES

1 Introduction

La 1^{ère} partie de cette thèse a permis d'évaluer et de comprendre l'importance du choix des modèles, des espèces et des variables de milieu pour expliquer et prédire la distribution actuelle des espèces. Cette distribution actuelle simulée est basée sur l'aptitude du modèle à décrire la distribution observée comme une combinaison des variables de milieu sélectionnées. Il est aisé ensuite de vérifier la fiabilité des prédictions en comparant la distribution observée avec celle simulée à l'aide de statistiques telles que la courbe ROC ou le Kappa. Projeter les distributions dans le futur repose sur cette combinaison de variables, déterminée par les modèles pour la situation actuelle, mais extrapolée pour les conditions futures. Parmi les variables de milieu couramment utilisées, les seules pour lesquelles des projections futures ont été réalisées sont les variables climatiques et les variables d'utilisation des terres et de couverture du sol.

Cela fait plus d'une décennie que des modèles climatiques ont été construits pour projeter dans un futur proche le climat pour 2020, 2050, voire 2100. Ces prévisions sont basées sur des Modèles de Circulation Générale (GCM). Ceux-ci sont des représentations des processus atmosphériques, des océans et des surfaces de terres et de glaces, basés sur des lois physiques connues (Hulme & Carter 2000). Depuis quelques années, ces modèles couplent les relations biosphère-atmosphère. Les GCM sont des outils qui estiment les changements de climat dus à l'augmentation des gaz à effet de serre, pour un grand nombre de variables de climat. Un inconvénient majeur dans l'utilisation de GCM est que, bien, qu'ils représentent le climat global d'une façon satisfaisante, leurs simulations du climat régional actuel sont souvent imprécises. Bien que les variables, à l'intérieur d'un GCM, soient toutes déterminées en utilisant des lois physiques ou des relations empiriques basées sur des lois physiques, les études de validation montrent que les relations internes entre les variables du modèle peuvent ne pas être nécessairement les mêmes que les relations observées dans le monde réel (Hulme & Carter 2000). Il en résulte que les GCM peuvent significativement sous-estimer ou surestimer les températures et les précipitations actuelles. Bien que les GCM aient des limitations pour prédire les scénarios de changements du climat actuel, ils fournissent la meilleure information disponible sur comment le climat global et régional peut changer suivant l'augmentation des concentrations des gaz à effet de serre. Depuis une dizaine d'années, le GIEC (Groupe d'experts Intergouvernemental sur l'Evolution du Climat) a demandé un gros effort à la communauté scientifique pour ajouter à ces modèles, différentes

Encadré II.1 Les différents scénarios d'émission et leurs caractéristiques (Watson et.al 2001).



A1FI, A1T et A1B

Le canevas et la famille de scénarios A1 décrivent un monde futur dans lequel la croissance économique sera très rapide, la population mondiale culminera au milieu du siècle pour décliner par la suite, et de nouvelles technologies plus efficaces seront introduites rapidement. Les principaux thèmes sous-jacents sont la convergence entre les régions, le renforcement des capacités et des

interactions culturelles et sociales accrues, avec une réduction substantielle des différences régionales dans le revenu par habitant. La famille de scénarios A1 se scinde en trois groupes qui décrivent des directions possibles de l'évolution technologique dans le système énergétique. Les trois groupes A1 se distinguent par leur accent technologique : forte intensité de combustibles fossiles

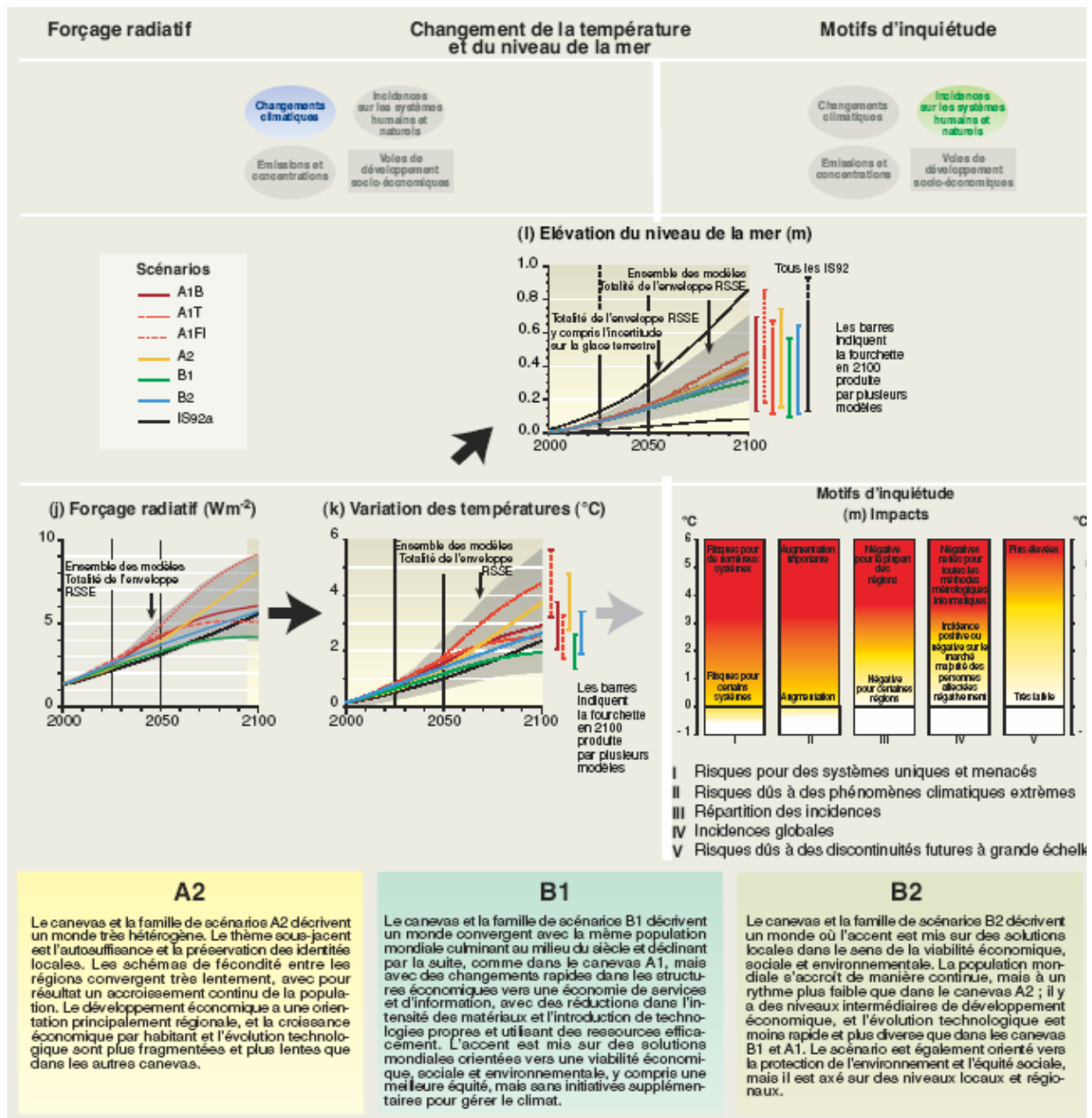
(A1FI), sources d'énergie autres que fossiles (A1T) ou équilibré entre les sources (A1B) (dans lequel « équilibré » est défini comme ne s'appuyant pas excessivement sur une source d'énergie particulière, en supposant que des taux d'amélioration similaires s'appliquent à toutes les technologies de l'approvisionnement énergétique et des utilisations finales).

variantes selon la réponse de la société aux changements climatiques (IPCC 1996). Ces variantes permettent pour un GCM donné, d'avoir des prévisions selon la croissance économique, le développement technologique ou la croissance de la population mondiale (Nakicenovic & Swart (2000)(Encadré II.1).

Les modèles de changements d'utilisation des terres et de couverture du sol sont plus récents et plus expérimentaux que les GCM. Ils prennent en compte l'occupation actuelle des terres, et avec des règles logiques comprenant des prédictions de la politique agricole commune, de la politique européenne, de l'économie et d'autres paramètres socio-économiques (par exemple le comportement des agriculteurs), prédisent les probables changements dans l'utilisation et l'occupation des terres. Il faut noter que ces modèles sont directement reliés aussi aux prédictions climatiques (Parry 2000).

Les modèles basés sur les niches utilisent donc ces scénarios de changements climatiques et d'utilisation des terres pour projeter les habitats potentiels des organismes modélisés. Ici, une très forte distinction est faite entre les futures distributions et les futurs habitats potentiels des organismes modélisés. Les modèles statiques fonctionnent sous le postulat très fort que les espèces sont en quasi-équilibre avec leur milieu. Les modèles calibrent donc les combinaisons des variables susceptibles d'expliquer et surtout de prédire cette distribution à l'équilibre, sans tenir compte de la dynamique intrinsèque des populations (dispersion ou compétition). Quand ces modèles statiques sont utilisés dans le futur, c'est la représentation spatiale des combinaisons de milieu représentant la distribution de l'organisme et non sa distribution elle-même qui est projeté. Autrement dit, les modèles projettent les zones où les combinaisons de variables sont adéquates pour l'organisme dans le futur, donc les habitats potentiels. Tant que les paramètres démographiques et dynamiques ne sont pas pris en compte, en aucune manière les futurs habitats potentiels d'une espèce peuvent être assimilés à sa future distribution. Il faudrait pour cela que l'espèce puisse rejoindre ces nouveaux habitats et s'y établir. La dispersion et la capacité de colonisation sont deux facteurs primordiaux, non pris en compte à l'heure actuelle dans de tels modèles compte tenu des lacunes persistantes dans les connaissances sur la dispersion et la capacité des espèces de suivre le déplacement géographique de leurs habitats favorables.

Dans cette seconde partie, je présenterai des développements méthodologiques et conceptuels permettant de mieux comprendre comment projeter les futurs habitats potentiels des espèces. La première étude traitera de l'aptitude des modèles à projeter les futurs habitats d'un grand nombre d'espèces. En effet, est-ce que deux modèles qualitativement similaires



pour prédire la distribution actuelle d'une espèce vont projeter des habitats similaires dans le futur ? Doit-on utiliser un seul et unique modèle (ex : GAM) pour prédire la distribution actuelle et l'habitat potentiel d'une espèce, ou alors choisir parmi une gamme de modèles (ex : GAM, GLM ou CTA) celui qui prédit le plus exactement la prédiction actuelle ? Même si le choix du modèle peut s'avérer prépondérant sur la projection des futurs habitats potentiels, il faut que celui-ci puisse faire une approximation convenable de la niche réalisée actuelle de l'espèce. Une étude sera donc développée pour montrer quelles peuvent être les répercussions de ne pas capturer convenablement la niche réalisée actuelle des espèces dans le but de projeter les futurs habitats potentiels dans de nouvelles conditions de milieu. Enfin, bien que j'aie démontré l'importance primordiale des variables macro-climatiques pour la prédiction des distributions des espèces à grande échelle, je présenterai une manière alternative de prendre en compte l'utilisation des terres et la couverture du sol dans le futur pour projeter les futurs habitats potentiels d'espèces de plantes, avec comme illustration une étude réalisée en Afrique du Sud.

2 Optimisation des prédictions et fiabilité des projections

Basé sur Annexe 2.1. :

Thuiller, W. 2003. BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9: 1353-1362.

De nombreuses études ont étudié et comparé les différents modèles existant sur leur capacité à prédire la distribution actuelle des espèces (Austin & Meyers 1996; Franklin 1998; Guisan *et al.* 1999 ; Manel *et al.* 1999; Vayssières *et al.* 2000; Thuiller *et al.* 2003a). Parmi toutes ces études, aucune n'a vraiment démontré qu'un modèle offrait tous les avantages nécessaires, indépendamment des espèces, des variables environnementales et de l'échelle. Chaque modèle a des caractéristiques propres alliant des avantages et malheureusement des inconvénients. Dans une optique de prédiction de la distribution d'un grand nombre d'espèces, on peut douter du consensus général d'utiliser un seul modèle pour toutes les espèces, surtout que le meilleur modèle va forcément varier entre toutes les espèces à modéliser. Bien souvent, les chercheurs utilisent le modèle qu'ils connaissent le mieux. A l'inverse, ne vaudrait-il pas mieux utiliser un modèle selon son aptitude à modéliser chaque espèce ?

Pour répondre à ces questions, j'ai développé un outil qui a pour nom BIOMOD (BIODiversity MODelling). J'ai programmé BIOMOD dans une librairie sous le logiciel Splus2000 (Anon 1999) en utilisant le langage orienté objet S (environ 3000 lignes de scripts).

Encadré II.2

Données espèces : 61 espèces de ligneux hauts ont été étudiées représentant les familles les plus importantes en Europe : Pinales, Taxales et Gnetales (Gymnosperme), Myricales, Juglandales et Fagales (Angiosperme). Les arbres ont été choisis car (i) leur écologie et distribution sont relativement bien connues comparé à d'autre taxa, (ii) leur richesse est corrélée (Corrélation de Spearman $\rho=0,80$; $P<0.001$) à la richesse totale de l'AFE (Araújo & Williams 2000) et (iii) ce sont des organismes à longue durée de vie et leur distribution est relativement stable comparé à d'autres groupes. Les données proviennent de l'AFE à 50x50 km.

Données de milieu : Seulement les sept variables bioclimatiques utilisées dans l'encadré I.4 ont été utilisées à la résolution de l'AFE pour calibrer les modèles. Ensuite les distributions ont été projetées à une résolution plus fine 10' correspondant à la résolution du projet ATEAM.

Scénario de changements climatiques : Le scénario utilisé pour 2050 provient du modèle de circulation global HadCM3 développé au Hadley Centre for Climate Prediction and Research et utilisant le scénario d'émission A1 (Cf. Encadré II.1).

Calibration de BIOMOD : La base de données a été divisée aléatoirement en une base de calibration (70%) utilisée pour calibrer les modèles et une d'évaluation (30%) utilisée pour évaluer la puissance prédictive des modèles.

4 types de modèles ont été paramétrés : Un GLM (avec termes linéaires, quadratiques et polynomiaux), un GAM (avec fonction de lissage '*Smooth Splines*'), un CTA et un ANN.

Evaluation de BIOMOD : La puissance prédictive des modèles a été évaluée à l'aide de deux méthodes : la statistique du Kappa et la courbe ROC. Cette dernière ne nécessite pas le calcul d'un seuil pour transformer les valeurs de probabilité en présence-absence, tandis que le Kappa a été calculé en utilisant un seuil maximisant cette statistique.

Prédiction et projection de BIOMOD : Chaque modèle a été calibré sur chaque espèce, comparé et utilisé pour dériver les prédictions spatiales sur les données originales (50x50 km). Ensuite, les modèles ont été utilisés pour projeter les distributions des espèces à une échelle plus fine (10'). Les futurs habitats potentiels de chaque espèce ont ensuite été dérivés en utilisant le scénario HadCM3 A2. Pour transformer les valeurs de probabilité de chaque modèle, deux méthodes ont été utilisées : soit en utilisant le seuil maximisant la statistique du Kappa, soit en utilisant le seuil maximisant le % de présences et d'absences correctement prédites déterminés sur les données actuelles

Tableau II.1: Moyenne, minimum et maximum de AUC et Kappa pour les 4 modèles sur les 61 espèces, suivant les données utilisées. Cal : données calibration ; Eval : données évaluation ; Pred : données originales. Min, me et max sont respectivement les valeurs minimum, moyenne et maximum pour chaque statistique.

	AUC									Kappa								
	Cal			Eval			Pred			Cal			Eval			Pred		
	min	me	max	min	Me	max	min	me	max	min	me	max	min	mea	max	min	mean	max
GLM	0.82	0.95	0.99	0.82	0.94	1.00	0.82	0.95	0.99	0.39	0.69	0.88	0.36	0.66	0.86	0.38	0.68	0.87
GAM	0.84	0.96	0.99	0.84	0.94	1.00	0.84	0.95	0.99	0.35	0.70	0.88	0.40	0.66	0.95	0.38	0.68	0.88
CART	0.83	0.94	0.99	0.54	0.87	0.96	0.82	0.92	0.98	0.45	0.72	0.88	0.27	0.59	0.84	0.45	0.68	0.86
ANN	0.87	0.97	1.00	0.82	0.95	0.99	0.85	0.96	0.99	0.54	0.79	0.94	0.34	0.68	0.84	0.50	0.75	0.89

La nouveauté de cet outil est de permettre l'utilisation automatisée des quatre types de modèles les plus largement utilisés en modélisation statique (GLM, GAM, CTA et ANN) pour prédire la distribution d'un nombre illimité d'espèces, de pouvoir toujours de manière automatique évaluer la puissance prédictive des modèles pour chaque espèce suivant deux techniques (courbe ROC et Kappa), et de projeter les futurs habitats potentiels (selon différents scénarios de changements globaux) des espèces en utilisant tous les modèles ou en utilisant seulement celui qui présente la meilleure puissance prédictive sur les données d'évaluation, et ceci pour chaque espèce.

Dans cette étude test, BIOMOD a été appliqué à 61 espèces d'arbres en Europe (Encadré II.2). Pour chaque espèce, les quatre modèles sont calibrés, évalués à l'aide de deux techniques (courbe ROC et Kappa), et utilisés pour prédire les distributions actuelles des espèces à 50x50, et projeter les distributions actuelles et futures à une résolution de 10'.

Evaluation : En moyenne, les distributions prédites sur les données de calibration et d'évaluation montrent une haute concordance avec les distributions observées (Tableau II.1). Les conclusions sont similaires selon qu'on utilise la courbe ROC ou la statistique du Kappa. Les réseaux de neurones artificiels semblent avoir en moyenne la meilleure puissance prédictive, tandis que les arbres de classifications ont la moins bonne. En comparant les résultats entre calibration et évaluation, on peut noter la tendance de CTA à surajuster la distribution des espèces durant le processus de calibration (très haute AUC) et à prédire avec une faible puissance la distribution des espèces sur les données d'évaluation (basse AUC). ANN tend aussi à surajuster durant la calibration, mais les prédictions sur les données d'évaluation montrent une très bonne puissance prédictive.

Prédictions et projections : Les prédictions pour les 61 espèces révèlent des réponses qui sont hautement dépendantes des espèces d'une part, et du modèle utilisé d'autre part. Généralement, la tendance à surestimer de CTA est évidente, et pour un certain nombre d'espèces, GLM, GAM et ANN diffèrent notablement. Concernant les projections dans le futur, la tendance attendue d'expansion ou de déplacement vers le nord est apparente pour un grand nombre d'espèces mais essentiellement pour les espèces natives du nord de l'Europe (ex : *Betula nana*). Cependant, pour beaucoup d'espèces, les changements dans les distributions potentielles tendent à être multidirectionnels. Pour illustrer cette étude, 4 espèces ont été sélectionnées représentant différentes classes chorologiques en Europe : une espèce

Figure II.1: Distributions observées et prédites pour *Quercus petraea* suivant le modèle employé. La courbe ROC et la statistique du Kappa sont les suivantes : ANN (AUC=0.96, $k=0.79$), CTA (AUC=0.927, $k=0.76$), GAM (AUC=0.945, $k=0.74$) et GLM (AUC=0.939, $k=0.74$)

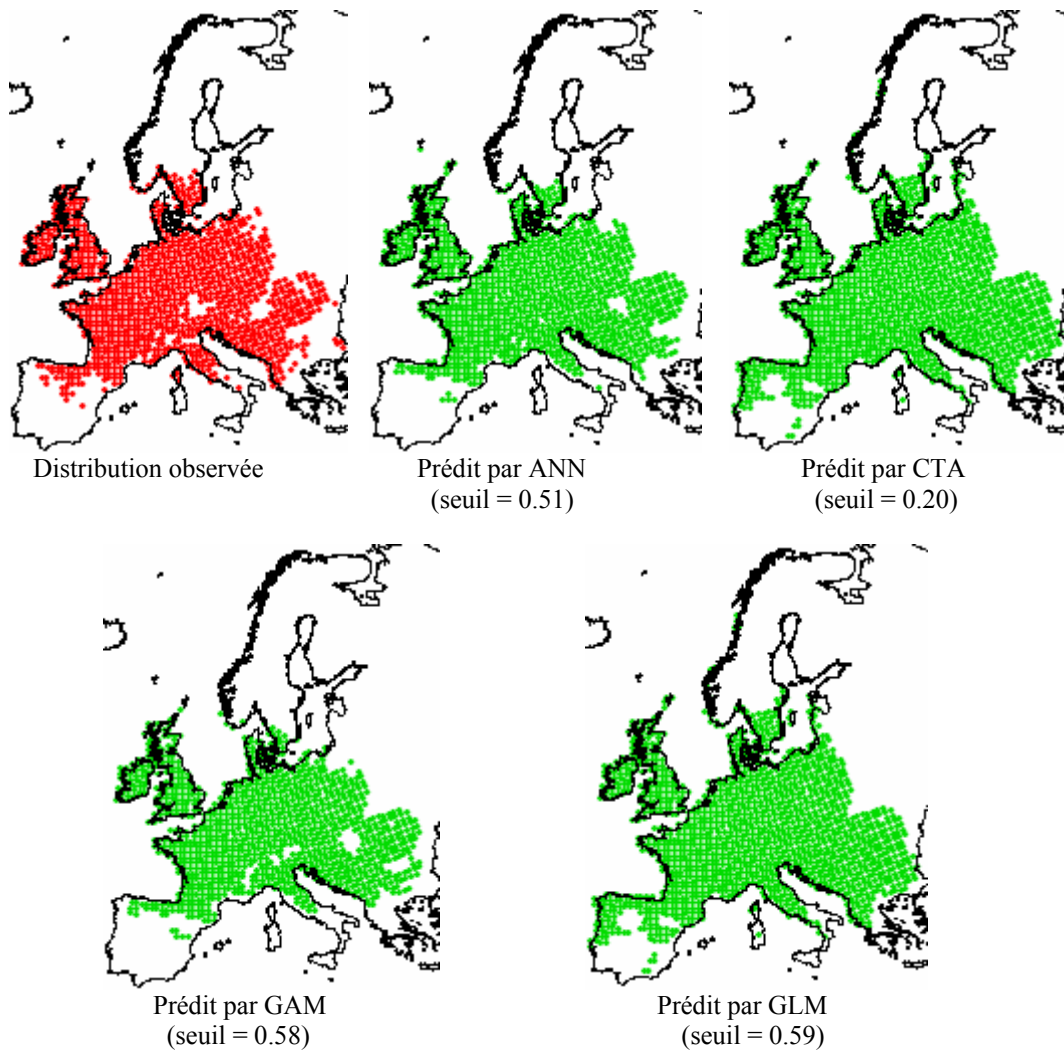
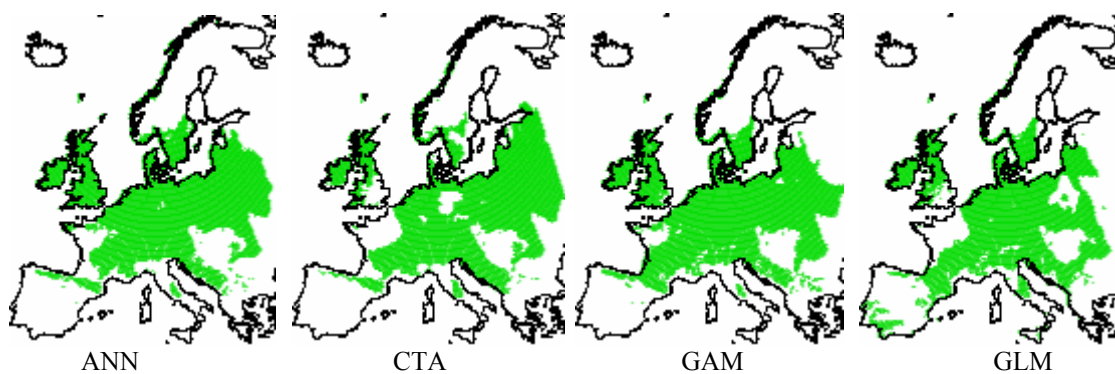


Figure II.2 : Futurs habitats potentiels de *Quercus petraea* projetés suivant le modèle employé. Les seuils de probabilité utilisés pour transformer les probabilités en données présence-absence sont les mêmes que pour la figure précédente.



eurosibérienne (*Quercus petraea*), une espèce continentale (*Castanea sativa*), une espèce méditerranéenne (*Pinus halepensis*) et une espèce sibérienne (*Betula nana*).

Les prédictions de la distribution actuelle de *Q. petraea* montrent des différences intéressantes suivant les modèles utilisés (Figure II.1). Premièrement, il y a une variabilité non négligeable dans les prédictions spatiales, même si AUC et Kappa montrent que les modèles ont la même puissance prédictive. Ceci indique que de petites différences en termes statistiques peuvent provoquer des différences notables en terme de prédictions spatiales.

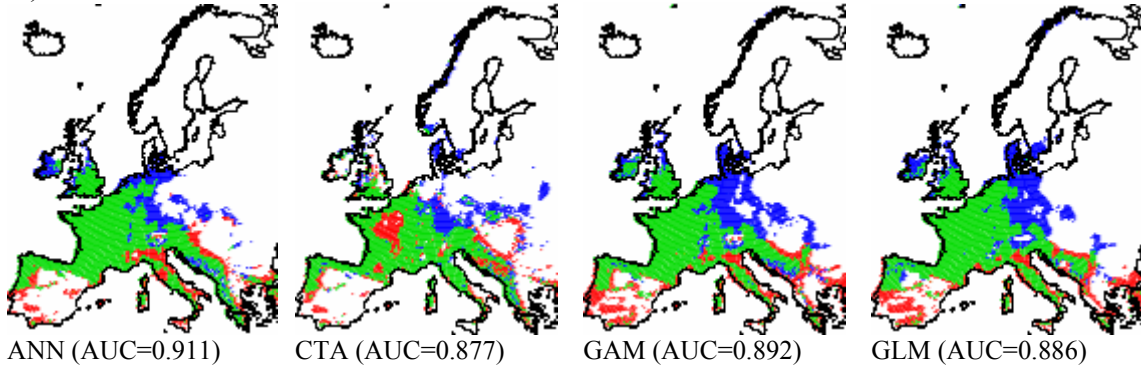
Les projections dans le futur pour cette même espèce montrent que de petites différences entre modèles pour les prédictions actuelles sont exacerbées dans le futur (Figure II.2). Selon les quatre modèles, la future distribution potentielle de *Q. petraea* pourrait s'étendre vers le nord de l'Europe, en admettant une dispersion et colonisation maximale. Cependant, CTA projette une distribution stable en Espagne et dans le sud de la France, alors que les trois autres modèles projettent une disparition des habitats adéquats dans ces zones. De plus, même si GLM, GAM et ANN montrent des projections similaires, il y a des différences notables en Grande-Bretagne, dans le nord de la France et en Pologne.

Les différences entre les futurs habitats potentiels des espèces selon la méthode utilisée ont une importance cruciale sur l'évaluation de la sensibilité des espèces aux changements climatiques. Par exemple, les quatre modèles diffèrent pour la projection des habitats perdus ou gagnés pour les espèces sélectionnées (Figure II.3-A). L'impact des changements climatiques sur *Castanea sativa* (Châtaigner) est différent selon le modèle utilisé. CTA projette une distribution stable de *C. sativa*, alors que les autres modèles projettent une expansion vers le nord et une contraction par le sud des habitats potentiels. Les résultats pour *Pinus halepensis* (Pin d'Alep) montrent aussi des divergences substantielles en fonction des modèles utilisés (Figure II.3-B). Pour cette espèce, il y a une grande différence entre les modèles statistiques et ceux basés sur les règles. GLM et GAM projettent une faible expansion vers le nord avec une contraction par le sud-est et par le sud-ouest, tandis que CTA et ANN projettent seulement une expansion vers le nord.

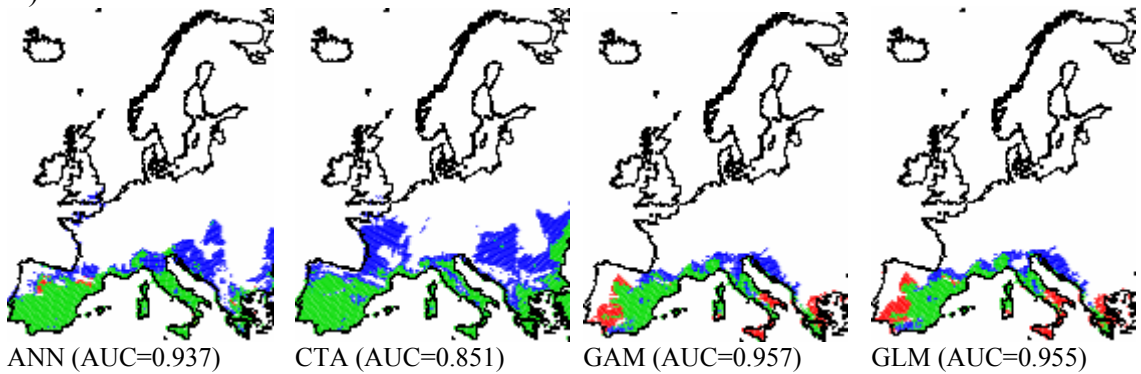
Cependant, quand les méthodes d'évaluation montrent une très faible différence entre modèles en terme de prédiction, la sensibilité des espèces est généralement la même entre les modèles. C'est le cas par exemple de *Betula nana*, une espèce sibérienne qui est projetée perdre une quantité considérable (40%) d'habitats climatiquement favorables (Figure II.3-C).

Figure II.3 : Comparaison des distributions actuelles projetées de *Castanea sativa* (A), *Pinus halepensis* (B) et *Betula nana* (C) et leurs projections dans le futur en 2050. Les couleurs verte, bleue et rouge correspondent respectivement aux habitats climatiquement adéquats actuellement et dans le futur ; aux habitats climatiquement adéquats dans le futur (pas actuellement) ; et aux habitats climatiquement adéquats actuellement et pas dans le futur.

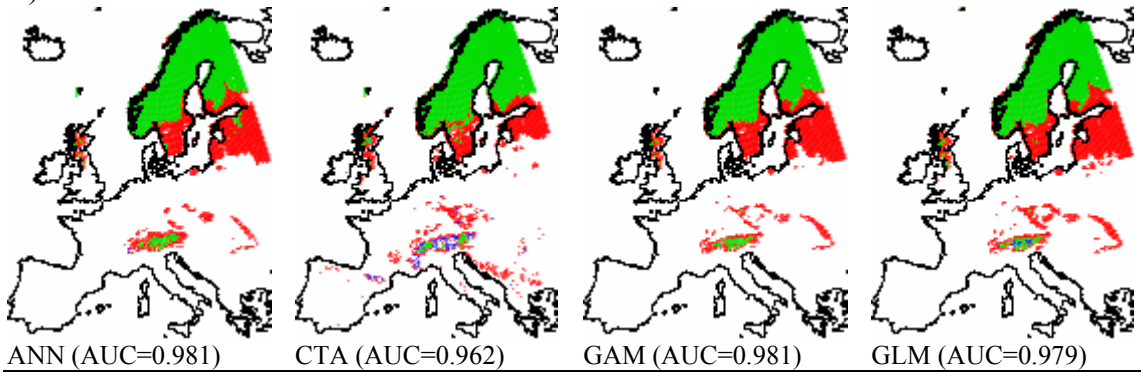
A)



B)



C)



Les résultats de cette étude suggèrent que les projections des futurs habitats potentiels des espèces dépendent fortement du modèle utilisé. Une petite différence entre deux modèles en termes de AUC ou Kappa peut résulter d'une variabilité importante dans les projections des futurs habitats potentiels. De plus, dans cette étude, je n'ai utilisé qu'un sous échantillon des formes possibles de ANN, CTA, GAM et GLM. Il y a d'autres façons de paramétrer ces modèles, par exemple utiliser une fonction de lissage *loess* au lieu des *splines* dans les GAM, ou bien en jouant sur le nombre de couches cachées dans le réseau de neurones. Une analyse de sensibilité serait intéressante à réaliser pour déterminer l'importance de la paramétrisation. Quoiqu'il en soit, tout ceci prouve l'utilité d'un outil comme BIOMOD qui choisi non pas le meilleur modèle selon les critères subjectifs du modélisateur mais selon sa capacité à prédire. Cet outil pourrait également servir à faire des comparaisons de modèles sur un grand nombre d'espèces, même si ceci n'est pas sa fonction initiale. En effet, de nombreuses études ont comparé des modèles selon leur aptitude à prédire la distribution de quelques espèces (Bio *et al.* 1998; Franklin 1998; Guisan *et al.* 1999; Vayssières *et al.* 2000; Moisen & Frescino 2002), mais il s'avère qu'aucun modèle ne semble universellement meilleur. Même en utilisant trois séries de données à trois échelles et résolutions différentes, je n'ai pas fourni de preuves irréfutables qu'un modèle soit supérieur aux autres (Thuiller *et al.* 2003a). Ici, je propose une approche alternative qui est d'utiliser différents modèles à l'intérieur d'une même structure pour toutes les espèces sélectionnées et de choisir le meilleur selon l'espèce. D'autres modèles pourraient compléter cette structure, comme par exemple les algorithmes génétiques fortement utilisés par Peterson et Stockwell (Peterson 2001; Peterson *et al.* 2001; Peterson *et al.* 2002; Peterson 2003), pour couvrir un plus large spectre de techniques de modélisation.

Si le but de l'étude est seulement de produire des prédictions sur données actuelles, BIOMOD est utile car il sélectionne le modèle le plus fiable pour chaque espèce. En ce sens, les prédictions sont optimisées, en comparaisons avec des approches n'utilisant qu'un seul type de modèle *a priori*. Evidemment, pour quelques espèces, les différents modèles employés donneront la même prédiction et il aura donc été inutile de perdre du temps à calibrer différents modèles (ex : *Betula nana*).

Un autre avantage de BIOMOD concerne les projections des futurs habitats potentiels des espèces. En effet, il est impossible de savoir si les projections sont exactes ou non, ou si un modèle est réellement plus prédictif que l'autre. Deux alternatives sont alors possibles :

- Soit décider de suivre le processus d'évaluation établi sur les données actuelles, sélectionner le modèle le plus performant pour prédire la distribution actuelle, et projeter dans le futur.

- Soit utiliser toutes les prédictions et projections des différents modèles et dériver des conclusions par comparaison des résultats, basées sur la connaissance d'experts ou des données publiées. Une voie complémentaire serait l'analyse des courbes de réponses futures des espèces aux gradients de milieu pour identifier des patrons irréalistes générés par des techniques particulières (Cf. paragraphe suivant). Enfin, une alternative serait de garder pour chaque espèce, le modèle dont les projections dans le futur résument l'information la plus pertinente de tous les modèles. Ceci permettrait d'avoir une approche conservative et consensuelle. Cette voie alternative sera développée dans la partie suivante.

Toutefois, la meilleure approche pour évaluer la performance des modèles dans le futur, serait de produire des projections rétrospectives de distributions d'espèces durant la dernière déglaciation. Les données climatiques pour ces périodes existent, de même que des données polliniques permettraient de comparer les projections aux cartes réelles de distribution (Harrison & Prentice 2003).

Pour finir, les projections des futurs habitats potentiels ont besoin d'être interprétées avec précaution. Si un modèle est utilisé pour projeter en dehors des bornes sur lequel il a été calibré, les projections peuvent être complètement biaisées (Thuiller *et al.* in press-b). Même si les modèles présentés ici sont assez puissants et fréquemment utilisés pour évaluer l'impact des changements globaux (Sykes *et al.* 2001; Bakkenes *et al.* 2002; Pearson *et al.* 2002), une attention particulière doit être portée sur les espèces en marge de distributions dans la zone d'étude et pour lesquelles les modèles ne peuvent capter l'enveloppe climatique entière (Thuiller *et al.* in press-b).

3 Approximation de la niche réalisée des espèces : conséquence pour les projections sous changements globaux

Basé sur Annexe 2.2 :

Thuiller, W., Brotons, L., Araújo, M. B & Lavorel, S. (In press). The effect of restricting the environmental range of data for predicting future species distributions. *Ecography*

On a vu que BIOMOD permettait d'optimiser les capacités prédictives des modèles pour une espèce donnée et donc d'augmenter la fiabilité des modèles. Cependant, ce type de structure automatisée n'attache pas beaucoup d'importance à la structure géographique des distributions d'espèces. En effet, quand on cherche à prédire la distribution actuelle et future d'un grand nombre d'espèces, il est difficile d'attacher de l'importance à chaque espèce particulière.

Encadré II.3

Données espèces : les trois distributions proviennent de l'Atlas Florae Europaea (Jalas & Suominen 1972-1996) à 50x50 km.

Données de milieu : Pour faciliter les résultats, nous avons utilisé seulement quatre variables censées avoir un impact sur les fonctions physiologiques des espèces ligneuses (Prentice *et al.* 1992) : ratio évapotranspiration réelle sur potentielle, moyenne des degrés-jours supérieurs à 5° (GDD5), précipitation annuelle et moyenne des températures du mois le plus froid. Les variables à 50x50 km ont été utilisées pour calibrer le modèle. Puis les variables à 10° ont été utilisées pour prédire les distributions futures.

Modèles : Trois modèles ont été calibrés pour chaque espèce : 1) en utilisant toutes les conditions bioclimatiques disponibles pour l'Europe (= 'référence'); 2) en utilisant les données bioclimatiques restreintes en limitant à 100 le nombre d'observations d'absences en dessous et en dessus de la dernière présence (= 'restreintes + 100') et 3) en utilisant les données bioclimatiques restreintes sans zéros au delà de la niche bioclimatique des espèces (=données restreintes). Le premier cas correspond à ce qui est communément utilisé pour projeter la future distribution des espèces (Bakkenes *et al.* 2002; Pearson *et al.* 2002; Thuiller 2003). Nous considérons ces données comme le modèle référence pour la comparaison avec les autres modèles. Le second cas permet de tester si l'inclusion d'un grand nombre d'absences en dehors de l'intervalle bioclimatique de l'espèce perturbe l'estimation des courbes de réponse et donc de la niche réalisée des espèces. Le troisième cas a pour but d'analyser l'impact de ne pas capturer la niche bioclimatique entière de l'espèce et de projeter les distributions futures en dehors de l'intervalle bioclimatique utilisée pour calibrer les modèles. Nous supposons que couper les données après la dernière présence revient à avoir une espèce en marge de distribution dans la zone d'étude ou avoir un nombre insuffisant de points d'échantillonnage.

Pour comparer la puissance prédictive des modèles, nous avons utilisé la courbe ROC réalisée sur les données d'évaluation (30% des données originales). Pour que les évaluations soient comparables, nous avons évalué les trois modèles sur les données d'évaluation des données restreintes.

Courbes de réponses projetées : Les courbes de réponses ont été utilisées pour projeter les distributions des espèces à une résolution plus fine (10')(Araújo *et al.* in press). Les projections pour le modèle référence ont été réalisées simplement en utilisant le modèle calibré sur toutes les données et en projetant sur les nouvelles données à 10'. Pour les données restreintes + 100, nous avons suivi la méthode d'Austin & Meyers (1996), en projetant les modèles seulement à l'intérieur de l'intervalle bioclimatique utilisée pour calibrer le modèle, et en mettant une probabilité égale à 0 pour les données bioclimatiques en dehors de cette intervalle. Supposant que le modèle 'restreint' pourrait correspondre à un sous échantillonnage de l'enveloppe bioclimatique complète de l'espèce, nous avons extrapolé la courbe de réponse sur les données de milieu à 10° sans aucune restriction. Alternativement, comme la méthode précédente peut mener à des distributions peu fiables, nous avons aussi projeté en suivant la méthode d'Austin & Meyers (1996), c'est à dire seulement à l'intérieur de l'enveloppe bioclimatique utilisée pour calibrer les modèles.

Cartographie des projections de distributions actuelles et futures à 10' : Nous avons ensuite cartographié les distributions actuelles et futures des trois espèces à partir de chaque modèle. Les distributions futures ont été générées en utilisant le modèle GCM HadCM3 avec le scénario A2 (voir Encadré II.1).

Tableau II.2 : Puissance prédictive des trois types de modèles projetés sur les données d'évaluations 'restreinte'. *Salix app* est l'abréviation pour *Salix appendiculata*. EcT correspond à l'écart type.

	Modèles	AUC	AUC
		Evaluation	EcT
<i>Pinus mugo</i>	"référence"	0.87	0.027
	"restreint + 100"	0.87	0.026
	"restreint"	0.86	0.027
<i>Salix app.</i>	"référence"	0.90	0.022
	"restreint + 100"	0.91	0.022
	"restreint"	0.91	0.022
<i>Quercus crenata</i>	"référence"	0.86	0.074
	"restreint + 100"	0.87	0.070
	"restreint"	0.86	0.063

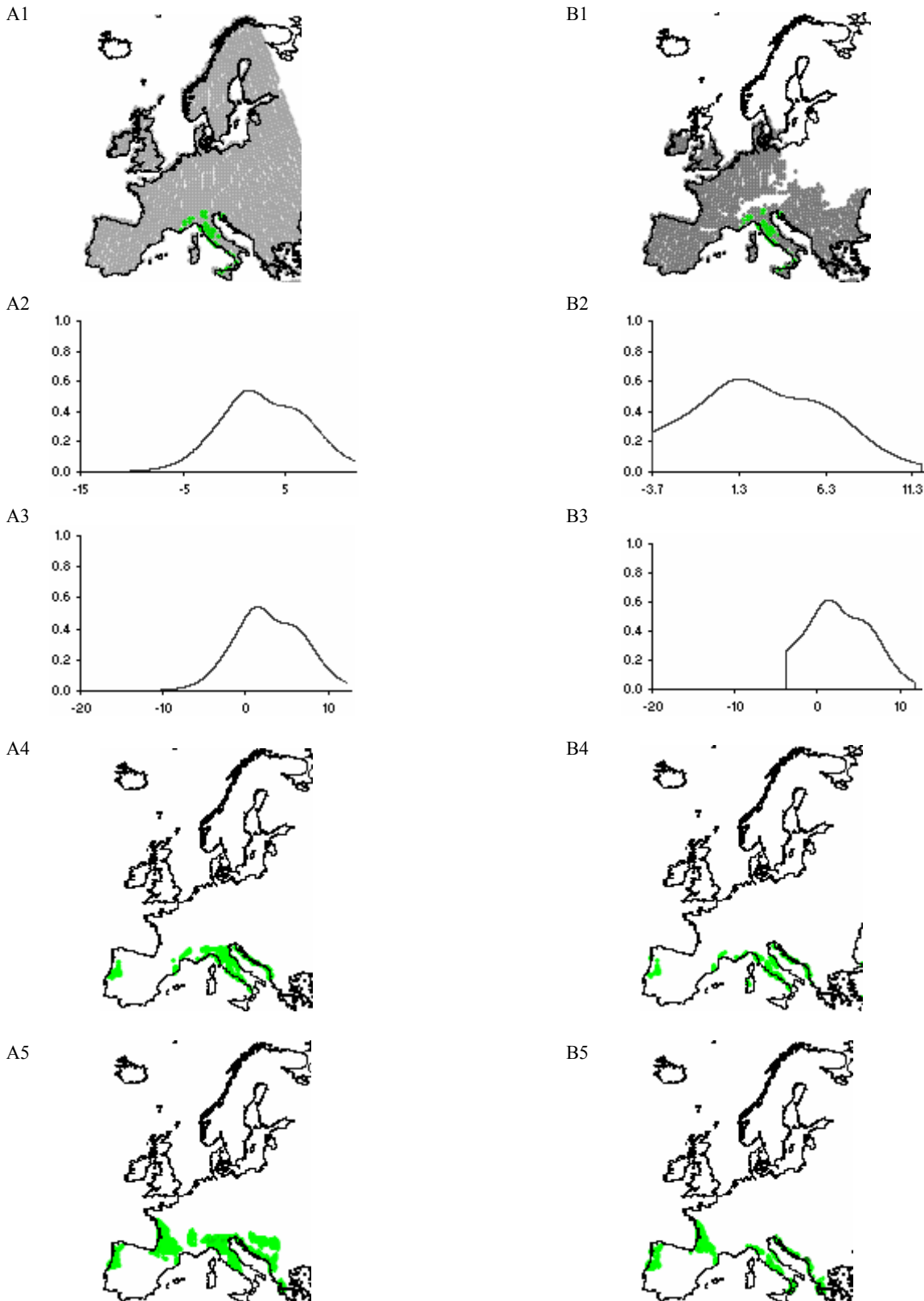
Les modèles statiques sont basés sur l'estimation de la niche réalisée des espèces. Si la niche réalisée d'une espèce est parfaitement modélisée, la représentation géographique sera très proche de la réalité. Pour qu'un modèle puisse projeter les futurs habitats potentiels d'une espèce d'une manière fiable, il lui faut modéliser au mieux sa niche réalisée bioclimatique actuelle (décrite par des variables bioclimatiques). La niche réalisée d'une espèce est communément approchée par une série de courbes décrivant sa probabilité d'occurrence dans la limite environnementale définie pour l'espèce. Un des problèmes de cette procédure, est que les modélisateurs manquent souvent d'informations sur la distribution potentielle ou réalisée des espèces. En effet, les courbes de réponse des espèces dans la zone d'étude sont souvent des descriptions incomplètes de la réponse des espèces aux variables de milieu. Différents auteurs ont suggéré des approches pour remédier à ce problème. Austin & Meyers (1996) débattent que l'inclusion de zéros (absences) situés en dehors de la limite bioclimatique (*'naughty noughts'*) d'une espèce perturbe l'estimation des courbes de réponse et donc la prédiction de la distribution. Ils suggèrent de restreindre les données aux limites environnementales de chaque espèce. Comme les modèles statistiques peuvent ajuster des zéros extrêmes raisonnablement, d'autres auteurs ont montré que cette restriction n'était pas nécessaire et pouvait mener à des courbes de réponse irréalistes (Oksanen & Minchin 2002).

De plus, les restrictions dues à la disponibilité des données de milieu sont souvent rencontrées pour la modélisation de distributions d'espèces à de grandes échelles. Ceci est fortement dû aux restrictions dans l'accès aux données et aux stratégies d'échantillonnage et aux contraintes géographiques (espèces en marge de son aire de distribution), empêchant la capture complète de la niche environnementale des espèces (Thuiller *et al.* 2003b).

Nous avons donc réalisé une étude pour analyser si effectivement l'inclusion d'un grand nombre de zéros (absences) situés en dehors de la limite bioclimatique d'une espèce trouble la bonne estimation des courbes de réponse et comment ceci modifie les projections futures. Cette étude a aussi pour but d'analyser les conséquences pour la qualité des prédictions de ne pas capturer la niche réalisée complète d'une espèce en marge de distribution et donc de n'obtenir que des courbes de réponse tronquées aux variables de milieu.

Comme présenté dans la partie précédente, peu de modèles permettent d'évaluer correctement les courbes de réponses des espèces aux variables de milieu. Certains supposent obligatoirement des courbes unimodales (ex : CCA) ou tandis que pour d'autres il est impossible de les extraire (ex : ANN ou GA). Pour étudier ce problème, nous avons donc choisis d'utiliser des modèles additifs généralisés (GAM) car leur qualité première est leur

Figure II.4 : Effets de la restriction de l'enveloppe des conditions climatiques utilisées pour calibrer les modèles sur la prédiction de *Quercus crenata* suivant la méthode conseillée par Austin et Meyers (1996). A gauche, modèle calibré sur toute l'Europe (A1) et la courbe de réponse actuelle de l'espèce à la T° du mois le plus froid (A2). Représentation de la courbe de réponse projetée sur les données actuelles à 10° (A3) et la carte projetée associée (A4). A5) Carte des habitats potentiels suivant cette procédure. A droite, modèle calibré sur les limites bioclimatiques de l'espèce en conservant 100 absences au delà de ses limites (B1). B2) Courbe de réponse calibrée à la T° du mois le plus froid ; B3) courbe de réponse projetée sur les données à 10° en limitant les prédictions sur les limites bioclimatiques utilisées (zéro au dehors) ; B4) Projection associée à B3 sur données actuelles et futures (B5).



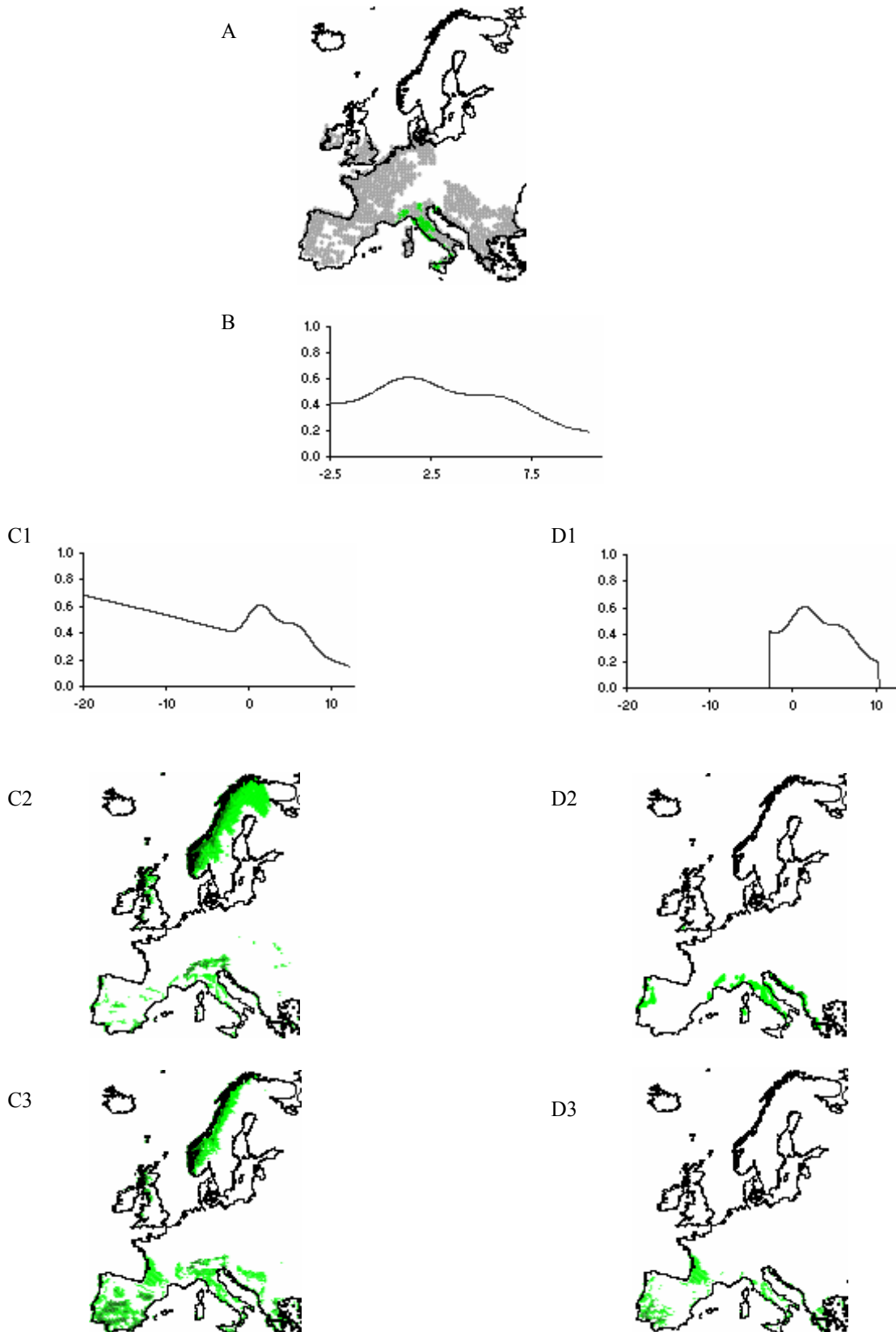
faculté d'ajustement des courbes de réponse (Austin *et al.* 1994; Austin & Meyers 1996; Oksanen & Minchin 2002). Ils démontrent de plus une très bonne puissance prédictive à tout type d'échelle et de résolution, sans pour autant être toujours les meilleurs (Thuiller *et al.* 2003a). Nous avons modélisé les courbes de réponse, les niches réalisées et les distributions actuelles et futurs habitats potentiels de trois espèces ligneuses européennes : *Quercus crenata*, *Pinus mugo* et *Salix appendiculata*. Ces espèces, avec des distributions restreintes et centrales en Europe, ont été choisies pour maximiser la probabilité d'inclure leur niche réalisée complète (Encadré II.3).

3.1 Effets de la restriction de l'enveloppe des conditions climatiques sur la fiabilité des modèles

Nous avons appliqué l'approche proposé par Austin & Meyers (1996) concernant les effets d'un grand nombre de zéros sur l'approximation des niches réalisées des espèces dans un autre contexte, qui est celui de la projection en dehors des limites géographiques utilisées pour calibrer les modèles et celui de la modélisation des distributions futures sous changements climatiques (Encadré II.3).

En terme de puissance prédictive, les modèles 'référence' et 'restreinte + 100' sont assez similaires, ce qui démontre qu'enlever des zéros ('naughty noughts') n'augmente pas la puissance prédictive des modèles (Tableau II.2). Les projections des courbes de réponse et niches réalisées sur la grille à 10' démontrent qu'enlever ces zéros en dehors de l'enveloppe bioclimatique de l'espèce procure une distribution plus conservative qu'en utilisant le modèle référence (Figure II.4). Ce résultat est en accord total avec Austin & Meyers (1996). Cependant, ces auteurs considèrent la sur-prédiction comme une erreur importante tandis que dans un contexte de changements climatiques nous tendons à estimer la sous-prédiction comme un problème plus grave. Dans le cadre de l'utilisation des modèles basés sur les niches pour projeter les distributions futures, les sur-prédictions raisonnables sont vues comme des habitats climatiquement adéquats pour l'espèce mais que celle-ci n'a pas encore colonisés pour cause de facteurs historiques ou de contraintes de dispersion. Du point de vue de la conservation, de tels habitats sont d'un grand intérêt car ils localisent des zones potentiellement adéquates, qui pourraient être utilisées pour la réintroduction d'espèces ou l'allocation de réserves. De plus, d'un point de vue écologique, les réponses des espèces au milieu peuvent être vues comme un continuum le long des gradients. Il semble donc plus intuitif que la probabilité de présence des espèces décroisse graduellement plutôt que de supposer une réponse tronquée.

Figure II.5 : Effet de ne pas capturer l'enveloppe climatique complète de l'espèce pour calibrer les modèles (ex : *Quercus crenata*). A) Modèle calibré sur les données restreintes et B) courbe de réponse à la T° du mois le plus froid. A gauche, Représentation de la courbe de réponse projetée sur les données actuelles sur toute l'Europe à 10° (C1) et la carte projetée associée (C2). C3) Carte des habitats potentiels suivant cette procédure. A droite, D1) courbe de réponse projetée sur les données à 10° en limitant les prédictions sur les limites bioclimatiques utilisées (zéro au dehors) ; D2) Projection associée à D1 sur données actuelles et futures (D3).



3.2 Effet de l'approximation incomplète de la niche bioclimatique des espèces

Restreindre complètement les données provoque plusieurs implications et conséquences :

Premièrement les absences sont souvent de vraies absences procurant une information relativement pertinente sur l'écologie des espèces. Utiliser des données restreintes (similaire à ne pas capter l'enveloppe bioclimatique complète de l'espèce) réduit considérablement les conditions de milieu sur lesquelles les modèles sont calibrés, et réduit l'applicabilité du modèle en terme de prédictions et projections (Figure II.5). Ce problème a d'importantes implications quand les projections dans le futur sont réalisées puisque certaines combinaisons de variables environnementales peuvent apparaître dans la région d'étude. En particulier, la modélisation de niches d'espèces provenant de données restreintes peut être vue comme analogue à la modélisation de niches d'espèces d'une zone géographique limitée ne couvrant pas l'enveloppe climatique complète des espèces. Dans ce cas, il y a deux possibilités pour la projection de la distribution de ces espèces sur de nouvelles conditions bioclimatiques (Figure II.5) : la première approche procure des projections très libérales, alors que la seconde sous-estime la vraie distribution. La question reste de savoir quelle est la meilleure approche pour modéliser et projeter dans le futur la niche de l'espèce à partir d'une zone limitée (espèce en marge de distribution) : Restreindre les projections à l'intervalle bioclimatique utilisé pour calibrer les modèles ? Ou projeter sans tenir compte d'aucune limite bioclimatique ? En l'absence d'informations précises, il peut être conseillé d'utiliser la méthode préconisée par Austin & Meyers (1996) pour projeter la distribution d'espèces en dehors de l'intervalle bioclimatique utilisée pour calibrer les modèles, et donc de restreindre les projections.

Deuxièmement, d'un point de vue technique, restreindre les conditions sur lesquelles les modèles sont calibrés peut générer des effets imprédictibles sur les queues des courbes de réponse (si les probabilités ne sont pas bornées à 0 en dehors des limites bioclimatiques utilisées pour calibrer le modèle), menant à des projections complètement imaginaires (Figure II.5). En effet, comme les données restreintes ne permettent pas la capture de la courbe de réponse dans son ensemble, les projections des distributions futures basées sur la niche écologique des espèces ne peuvent être valides car les modèles ne sont pas capables d'approcher la courbe de réponse complète des espèces sur les variables de milieu.

Comme exemple de comment considérer de telles contraintes, Pearson *et al.* (2002) réalisent des modèles basés sur les niches à échelle européenne pour augmenter la probabilité d'inclure les limites bioclimatiques complètes des espèces étudiées et ensuite utilisent ces modèles pour prédire la distribution des espèces en Grande Bretagne. Ceci assure que les

Encadré II.4

Données espèces : 28 espèces endémiques de la famille des Proteacées incluant les genres *Diastella*, *Leucadendron*, *Leucospermum*, *Protea* et *Serruria*. Toutes ces espèces sont dioïques, ligneuses, majoritairement des arbustes et occupant une large gamme de sols et d'altitude. La distribution spatiale de ces espèces provient d'un échantillonnage massif de toute la famille des *Proteacées* (Rebelo 1992). La présence des espèces a été notée sur plus de 30000 sites à travers la région et est disponible sous format SIG (Rebelo 2001).

Variables de milieu : Les variables utilisées sont la moyenne des températures minimum du mois le plus froid, les degrés-jours (GDD18°), l'évapotranspiration annuelle, l'humidité du sol en hiver et en été. Les deux dernières variables réfèrent au nombre de jours pour lesquels l'humidité du sol et la température de l'air sont favorables pour la croissance de la plante. Ces variables sont parmi celles dérivées pour l'Afrique du Sud pour le climat présent et futur en 2050 par Schulze (1997) et Schulze & Perks (1999).

Modèles : Des modèles GAM ont été utilisés sous une forme saturée, c'est à dire sans sélection de variables. Pour évaluer leur performance, la déviance expliquée des modèles a été calculée comme 1 - le ratio déviance résiduelle sur déviance nulle. Cette statistique est analogue au R^2 des modèles de régressions basés sur les moindres carrés. Si le modèle n'améliore pas le modèle nul (aléatoire) pour une espèce donnée, alors l'espèce est supposée ne pas avoir de limitations bioclimatiques ou édaphiques et sa réponse aux changements climatiques n'a pas été examinée (cas de deux espèces).

Impact de l'utilisation des terres : L'impact de l'utilisation des terres sur les futurs habitats potentiels des espèces a été évalué en supprimant les zones actuellement occupées par des pratiques ne permettant pas la croissance d'une espèce (CSIR 1999). De cette manière, deux types de cartes sont réalisées dans le futur : une sans prendre en compte l'utilisation des terres et l'autre en enlevant les zones où l'utilisation des terres pourrait empêcher le développement ou la croissance des espèces.

modèles ne sont pas utilisés pour extrapoler au delà des limites bioclimatiques utilisées pour calibrer les modèles.

4 Futurs habitats potentiels des espèces : la prise en compte de l'utilisation des terres *a posteriori*

Basé sur Annexe 2.3 :

Midgley, G. F., Hannah, L., Millar, D., Thuiller, W. and Booth, A. (2003). Developing regional and species-level assessments of climate change impacts on biodiversity: A preliminary study in the Cape Floristic Region. *Biological Conservation*, 112, 87-97.

Dans la première partie du manuscrit, j'ai démontré que l'utilisation des terres et couverture du sol n'étaient pas des variables puissantes pour prédire la distribution des espèces à grande échelle (Thuiller *et al.* in press-a). Cependant, l'intensification de l'agriculture dans certaines régions ou la déprise agricole dans d'autres sont potentiellement des facteurs influençant, en bien ou en mal, la distribution actuelle des espèces et encore plus probablement leurs distributions futures (Mooney & Hobbs 2000; Robinson *et al.* 2001; Robinson & Sutherland 2002). En effet à échelle plus locale, l'utilisation des terres et la couverture du sol (UT/CS) influencent la distribution des espèces de différentes manières, que ce soit par l'augmentation de la fragmentation du paysage (Robinson *et al.* 1992), ou l'homogénéisation des paysages (Robinson *et al.* 2001). Outre la détermination du meilleur modèle pour projeter les futurs habitats potentiels des espèces (ex : BIOMOD), ou l'attention portée à la capture complète des limites bioclimatiques des espèces, la prise en compte de l'UT/CS pourraient augmenter la fiabilité et la légitimité des projections. Dans une analyse réalisée en Afrique du Sud (en collaboration avec Guy Midgley et Lee Hannah pour Conservation International), nous avons donc cherché à inclure l'UT/CS dans des modèles prédictifs. Ces modèles ont pour but de projeter les impacts des changements climatiques dans la Région Floristique du Cap sur le biome (fynbos) bien particulier des *Proteacées*, en vue de proposer une sélection de réseaux de conservation pour atténuer les effets dévastateurs des changements globaux sur ces espèces, dont la plupart sont endémiques. Comme dans cette région, la transformation des terres est un facteur important, mais comme il n'existe pas à l'heure actuelle de scénarios de changements, nous avons pris en compte ce facteur *a posteriori*. En d'autres termes, nous avons projeté les futurs habitats potentiels des espèces de *Proteacées* sur la base d'un scénario de changements climatiques et analysé les zones où théoriquement ces habitats ne seraient pas viables compte tenu des pratiques humaines actuelles.

Tableau II.3 : Surface de l'aire de distribution modélisée actuelles et futures pour 28 espèces de Proteacées (=nombre total de cellules de 1x1 km avec une probabilité d'occurrence supérieure à 0.5), en prenant ou pas en compte l'utilisation des terres. UT signifie Utilisation des Terres.

Espèces	Actuelle		2050	
	Sans UT	Avec UT	Sans UT	Avec UT
<i>Diastella buekii</i>	12	4	3	3
<i>Diastella parilis</i>	15	9	0	0
<i>Diastella proteoides</i>	63	26	114	53
<i>Leucadendron chamelaea</i>	45	16	1	1
<i>Leucadendron cinereum</i>	154	55	1317	538
<i>Leucadendron corymbosum</i>	66	35	5	3
<i>Leucadendron foedum</i>	1	1	83	73
<i>Leucadendron galpinii</i>	218	112	1	1
<i>Leucadendron lanigerum</i> var. <i>lanigerum</i>	352	71	43	24
<i>Leucadendron levisanus</i>	17	3	308	219
<i>Leucadendron stellare</i>	103	11	366	147
<i>Leucadendron thymifolium</i>	15	0	257	206
<i>Leucospermum arenarium</i>	27	10	5	0
<i>Leucospermum hypophyllocarpodendron</i>	263	114	30	19
<i>Leucospermum muirii</i>	45	26	0	0
<i>Leucospermum parile</i>	81	39	1	1
<i>Leucospermum rodolentum</i>	1276	707	3	3
<i>Leucospermum tomentosum</i>	513	319	372	216
<i>Protea scolymocephala</i>	2	0	35	24
<i>Serruria adscendens</i> var. <i>decipiens</i>	9	8	165	131
<i>Serruria brownii</i>	1	0	0	0
<i>Serruria candicans</i>	22	22	9	9
<i>Serruria cyanoides</i>	0	0	2	2
<i>Serruria decipiens</i>	149	78	0	0
<i>Serruria decumbens</i>	0	0	18	17
<i>Serruria fucifolia</i>	744	377	21	19
<i>Serruria linearis</i>	28	15	0	0
<i>Serruria trilopha</i>	0	0	108	102

Nous avons sélectionnée 28 espèces de Proteacées, toutes endémiques de la Région Floristique du Cap (CFR) permettant ainsi aux modèles de capturer leurs limites bioclimatiques complètes. La CFR devrait être fortement soumise à des variations climatiques, alors qu'elle possède une richesse floristique exceptionnelle avec un très fort taux d'endémisme (Cowling *et al.* 1989). Le climat de type méditerranéen est en train de changer, par une augmentation des températures et une diminution des précipitations (Tyson *et al.* 2002).

Pour modéliser la distribution actuelle des espèces de Proteacées et les habitats potentiels en 2050, nous avons utilisé cinq variables bioclimatiques considérées comme critiques pour les fonctions physiologiques et la survie des plantes (Woodward 1987; Woodward & Williams 1987). Comme il s'agit plus d'une analyse préliminaire que d'une réelle tentative de projeter les futurs habitats potentiels, le choix du modèle a été de prendre une approche bien connue et qui a montré de bons résultats dans des analyses précédentes (Yee & Mitchell 1991; Leathwick *et al.* 1998; Bio *et al.* 2002; Thuiller *et al.* 2003a), les modèles GAM (Encadré II.4).

Les 28 espèces modélisées montrent une grande variabilité dans leur réponse (Tableau II.3), procurant de bonnes indications quant aux risques dus aux changements climatiques. Trois sortes de risques peuvent être ainsi extrapolés :

Disparition de l'aire de répartition : 5 des 28 espèces (18%) sont projetées disparaître entièrement sous changements climatiques. Ces espèces peuvent donc être considérées comme ayant un fort risque d'extinction dans le futur même si certaines précautions doivent être prises. Bien que basée sur une connaissance fiable des limites bioclimatiques des espèces, la perte modélisée de l'aire de distribution peut ne pas représenter une extinction immédiate. Il est possible par exemple, que des adultes établis persistent dans des régions en dehors des limites bioclimatiques de ces espèces, car les limites sont justement basées sur la sensibilité d'établissement des graines.

Diminution de l'aire de répartition : 12 des 28 espèces (42%) sont projetées subir une diminution de leur aire de distribution dans le futur. Parmi ces espèces, la perte moyenne est de 84% avec des extrêmes compris entre 27 et 99%.

Bien que la perte d'habitat prédomine, une proportion notable d'espèces (39%) montre un gain potentiel d'habitats adéquats. Ces résultats corroborent les résultats de Warren *et al.*

Tableau II.4 : Plus courte distance entre la distribution actuelle et la distribution future potentielle des espèces, en tenant compte ou non de l'utilisation des terres. Centroïde-centroïde signifie la distance entre le centroïde de la distribution actuelle et la future potentielle, et Limites-limites signifie la distance entre les limites de distributions les plus proches. Une distance de 0 signifie que la distribution actuelle de l'espèce et son habitat potentielle future se chevauche. UT signifie utilisation des terres. La distance est exprimée en kilomètre (km). E signifie extinction.

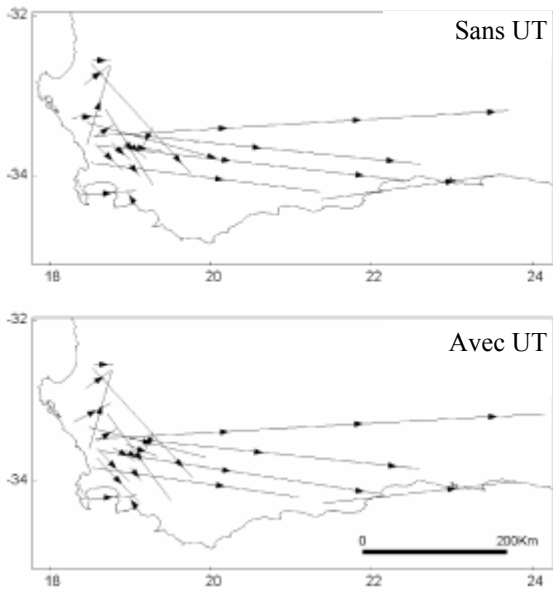
Espèces	Centroïde-centroïde (km)		Limites-limites (km)	
	Sans UT	Avec UT	Sans UT	Avec UT
<i>Diastella buekii</i>	22	22	15	15
<i>Diastella parilis</i>	E	E	E	E
<i>Diastella proteoides</i>	49	59	0	5
<i>Leucadendron chamelaea</i>	28	28	8	8
<i>Leucadendron cinereum</i>	154	142	0	0
<i>Leucadendron corymbosum</i>	36	40	0	0
<i>Leucadendron foedum</i>	40	40	0	0
<i>Leucadendron galpinii</i>	206	224	7	7
<i>Leucadendron lanigerum</i>	33	31	0	0
<i>Leucadendron levisanus</i>	267	239	0	3
<i>Leucadendron stellare</i>	119	141	8	17
<i>Leucadendron thymifolium</i>	357	328	11	38
<i>Leucospermum arenarium</i>	190	190	157	157
<i>Leucospermum hypophyllocarpodendron</i>	483	523	2	6
<i>Leucospermum muirii</i>	E	E	E	E
<i>Leucospermum parile</i>	29	29	22	22
<i>Leucospermum rodolentum</i>	10	10	5	5
<i>Leucospermum tomentosum</i>	35	50	0	0
<i>Protea scolymocephala</i>	111	111	21	21
<i>Serruria adscendens var. decipiens</i>	19	17	0	0
<i>Serruria brownii</i>	E	E	E	E
<i>Serruria candicans</i>	340	340	0	0
<i>Serruria cyanoides</i>	7	7	0	0
<i>Serruria decipiens</i>	E	E	E	E
<i>Serruria decumbens</i>	65	63	2	2
<i>Serruria fucifolia</i>	22	22	2	2
<i>Serruria linearis</i>	E	E	E	E
<i>Serruria trilopha</i>	101	80	5	22

(2001), Peterson *et al.* (2002) et Thuiller (2003) qui ont trouvé que les espèces développent des réponses très différentes aux changements climatiques, positives ou négatives, dépendant principalement de leur relation avec le milieu. Pimm (2001) a d'ailleurs noté que cette diversité de réponses pose des problèmes aux efforts de modélisation prédictive. Même si certaines espèces sont prédites étendre leur aire de distribution, ceci dépendra fortement de leur capacité de dispersion, de la taille de la population, de la fécondité et de leur capacités à coloniser de nouveaux habitats. L'information sur ces paramètres n'est pas disponible pour les espèces sélectionnées. De plus, la classification assez grossière de la fertilité et de la texture du sol que nous utilisons peut surestimer les habitats potentiellement adéquats pour les espèces, ce qui peut provoquer des prédictions optimistes dans le futur.

L'utilisation des terres a un impact négatif très clair sur les aires potentielles de distributions des *Proteacées* dans le futur. Sous les conditions climatiques actuelles, en moyenne 55% des zones potentielles pour les espèces ne sont pas utilisables car actuellement occupées ou transformées par l'homme. Ces résultats ne tiennent bien sûr pas compte de l'augmentation croissante de la modification des terres d'origines anthropiques.

Translations d'aire de répartition : 23 espèces parmi les 28 modélisées montrent un mouvement géographique de leur habitat potentiel. 13 de ces espèces ne montrent pas de chevauchement entre leurs distributions actuelles et les habitats projetés après la prise en compte de l'utilisation des terres (Tableau II.4). Les vecteurs des mouvements de distributions (dérivés à partir des centroïdes) suggèrent un déplacement général en direction du sud-est en réponse aux changements climatiques projetés (Figure II.6). Les vecteurs distances liant les distributions actuelles simulées et les habitats potentiels futurs sont compris entre 8 et 483 km (Tableau II.4). Cependant, la distance entre les centroïdes de distributions n'est peut-être pas la meilleure mesure pour les espèces avec des aires de distributions très vastes. En regardant les distances entre les plus proches limites des distributions (Tableau II.4), la séparation des aires de distributions est significativement inférieure à 25 km dans de nombreux cas, même en tenant compte de l'utilisation des terres. 10 espèces montrent un chevauchement entre leurs distributions actuelles et leurs distributions potentielles futures. Cependant, ce nombre passe à 8 en tenant compte de la modification d'utilisation des terres, ce qui montre que la destruction d'habitats augmente le risque dû aux changements climatiques, en accentuant le besoin de disperser à longue distance pour de nombreuses espèces. Pour finir, il est intéressant de noter que parmi les espèces qui sont projetées étendre leurs aires de distribution, la majorité font partie des espèces qui sont prédites avoir une translation complète de leurs distributions (pas

Figure II.6: Mouvement des aires de distributions des espèces modélisées. La norme des vecteurs correspond à la distance entre le centroïde de la distribution actuelle et celui de la distribution future. La flèche indique la direction du mouvement. Sans UT et Avec UT correspondent respectivement aux cas sans prise en compte de l'utilisation des terres et au cas avec prise en compte.



de chevauchement entre distributions actuelles et futures potentielles). Ces espèces doivent migrer sur des distances considérables pour éviter l'extinction, malgré le fait que les modèles aient projeté une augmentation de leurs aires de distributions. Ceci indique que l'augmentation de l'aire de distribution n'est pas un indicateur fiable de la sensibilité des espèces aux changements globaux. La modification de l'utilisation des terres sera un paramètre critique dans la détermination des liens paysagers pour ces espèces sans chevauchement de distribution.

En conclusion, si l'information est disponible, l'utilisation des terres, la couverture du sol et leur modification dans le futur, sont des paramètres qui peuvent être pris en compte à condition d'avoir une approche s'intéressant séparément au cas de chaque espèce. De plus, ceci reste valable avec des données à très fines résolutions (<1x1 km), mais l'est beaucoup moins à des résolutions plus grossières (50x50 km).

5 Conclusion

Cette partie avait pour but de mettre en évidence les approches, les avantages et les biais des modèles statiques dans un contexte de changements climatiques. Les modèles statiques cherchent à faire une approximation de la niche écologique des espèces, basée sur des combinaisons de variables de milieu. Ces combinaisons, qu'elles soient linéaires ou non, définissent des limites, dans un espace à n -dimensions, de probabilités d'existence des espèces. Ces limites sont ensuite extrapolées dans le futur par l'utilisation de scénarios de changements sur les variables de milieu utilisées pour calibrer les modèles. Ces scénarios de changements de milieu sont actuellement de deux types inter-corrélés : l'un porte sur le climat et l'autre sur l'utilisation des terres. Les scénarios de changements climatiques sont basés sur des modèles de circulation globale (GCM) qui modélisent les relations atmosphère-biosphère et tentent de projeter le climat futur en fonction de différentes prévisions de variables physiques comme les émissions de carbone ou de composés azotés.

Toutefois, projeter les futurs habitats potentiels pose certains problèmes d'ordre méthodologique et écologique auxquels j'ai tenté de répondre :

- 1- Il n'existe pas de modèle universellement meilleur que les autres et il est très difficile d'établir une hiérarchie simple et efficace. De plus, il est impossible d'estimer la fiabilité des modèles dans le futur. J'ai développé BIOMOD, une plateforme sous le logiciel Splus permettant de tester d'une manière automatisée pour chaque espèce une gamme de modèles statiques. Cet outil permet de choisir le meilleur modèle pour chaque espèce sur les conditions actuelles au lieu de choisir un modèle *a priori* pour toutes les espèces

étudiées. En utilisant BIOMOD sur 61 espèces, j'ai montré que des modèles avec de bonnes puissances prédictives sous des conditions actuelles pouvaient projeter des futurs habitats potentiels assez différents et donc provoquer des conclusions différentes. Il y a donc un intérêt majeur à tester différents modèles pour chaque espèce étudiée et examiner attentivement leurs projections plutôt que d'utiliser un seul modèle sans savoir s'il parviendra à estimer correctement la distribution actuelle et peut-être future des espèces. A ce stade, trois stratégies s'offrent pour projeter les futurs habitats potentiels : 1) Utiliser un seul modèle *a priori* pour toutes les espèces sélectionnées. Dans ce cas, la puissance prédictive n'est pas maximisée. 2) Utiliser BIOMOD ou autres structures du même type, testant et comparant une gamme prédéfinie de modèles sur les conditions actuelles et utiliser le meilleur pour projeter les futurs habitats potentiels. Le problème est de savoir si le meilleur modèle pour prédire la distribution actuelle d'une espèce sera le meilleur pour projeter ses futurs habitats potentiels. 3) Utiliser BIOMOD comme dans le cas précédent, mais effectuer les projections en utilisant tous les modèles, ce qui permet de donner un indice d'incertitude (où les modèles sont en désaccord) et un indice de fiabilité (où les modèles sont d'accord). De plus, une analyse plus fine du comportement des modèles (courbe de réponses) permet d'évaluer les modèles sur la base de critères écologiques et de leur faculté à ajuster les niches réalisées des espèces.

- 2- La faculté des modèles à ajuster les réponses des espèces aux variables de milieu et à faire une approximation de leur niche réalisée est un paramètre essentiel pour la fiabilité des projections dans le futur. Les modèles ont de la difficulté à modéliser les espèces en marge de distributions sur la zone d'étude, ce qui peut provoquer des erreurs importantes de sur- ou sous-prédictions. Les modèles ne pouvant ajuster convenablement la niche réalisée de ces espèces, perdent de la fiabilité quand les distributions sont projetées en dehors des limites des variables utilisées pour calibrer les modèles. Une alternative à cela serait de ne pas projeter de distribution au delà des limites de milieu utilisées pour calibrer les modèles, ce qui est aisé quand on étudie une seule espèce mais devient fastidieux quand on étudie un grand nombre d'espèces.
- 3- Une conclusion importante de la première partie était que la prise en compte de l'utilisation des terres pour prédire la distribution des espèces à grande échelle n'était pas nécessaire. Cependant, compte tenu de la forte pression exercée par le changement d'utilisation des terres à échelle plus locale, du manque de scénarios fiables d'utilisation des terres dans le futur et de la nécessité de connaître leurs effets, nous avons présenté une alternative intéressante qui permet de prendre en compte ce facteur *a posteriori*. Les futurs

habitats potentiels de 28 espèces de *Proteacées* en Afrique du Sud ont été projetés sous changements climatiques. Ces habitats ont ensuite été restreints aux zones non occupées par des cultures ou autres utilisations empêchant la colonisation de ces milieux. Bien que les changements d'utilisation des terres n'aient pas été pris en compte, cette étude montre qu'il est possible dans des études régionales d'utiliser ces variables *a posteriori* car elles offrent une bonne représentation de la disponibilité des terres pour la colonisation et la dispersion des espèces.

3^{EME} PARTIE

SENSIBILITES DES ESPECES ET SENSIBILITE

GEOGRAPHIQUE AUX CHANGEMENTS

CLIMATIQUES : PROJECTIONS ET INCERTITUDES

1 Introduction

Les deux premières parties de cette thèse ont permis de faire un état des lieux des concepts, des hypothèses et des méthodes qu'offre la biogéographie prédictive pour évaluer la sensibilité des espèces aux changements globaux. Il y a différentes mesures de sensibilité d'une espèce : par exemple la distance entre le centroïde de la distribution actuelle et celui de la distribution potentielle future (Midgley *et al.* 2003). Cette mesure présente quelques inconvénients car les modèles statiques ne prennent en compte ni les capacités de dispersion et de colonisation des espèces, ni la structuration du paysage. Une autre mesure très similaire à la précédente, est la perte ou le gain de pixels d'habitats favorables (Bakkenes *et al.* 2002; Thuiller 2003). Cette mesure offre des avantages et des inconvénients similaires à la précédente, mais permet de séparer les sites perdus, de ceux gagnés par chaque espèce. En effet, s'il est difficile actuellement d'estimer la fiabilité de nouveaux habitats potentiels pour une espèce sans connaître ses capacités de dispersion ni la structure du paysage, la fiabilité des habitats qui deviennent inadéquats peut sembler plus forte. Celle-ci dépend de la relation mesurée entre les variables de milieu déterminant la distribution de l'espèce et l'espèce elle-même. Toutefois, même si la rapidité exceptionnelle des changements climatiques permet de s'affranchir des processus évolutifs pour la plupart des espèces longévives, la plasticité phénotypique pourrait permettre à certaines espèces de perdurer dans des habitats prédits comme inadéquat. Cependant, la plasticité et des temps de réponses décalés sont des paramètres inconnus et très difficiles à estimer en fonction des futurs changements climatiques.

Les mesures précédentes de sensibilité sont indéniablement liées à la notion d'espèce individuelle, mais des mesures de sensibilités géographiques des pools régionaux d'espèces par pixel peuvent aussi être obtenues. En effet, les modèles permettent d'évaluer la richesse spécifique ou taxonomique actuelle par pixel (somme des présences d'espèces simulées par pixel) et d'estimer dans le futur quelle est la proportion d'espèces qui devraient potentiellement disparaître ou coloniser chaque pixel. Ces deux mesures permettent de déduire potentiellement le turnover d'espèces dans le pool régional présent dans chaque pixel. Plus le turnover spécifique est important, plus il pourrait avoir des changements d'espèces à l'intérieur du pool initial de chaque pixel. Même si la résolution à laquelle nous travaillons (50x50 km ou 10') ne nous permet pas explicitement d'étudier l'impact sur les écosystèmes ou sur les communautés, il est évident qu'un fort taux de turnover d'espèces dans chaque pixel provoquera une restructuration au niveau de la végétation, des changements probables dans la

Encadré III.1

Données espèces : Sous échantillon de l'AFE (Lahti & Lampinen 1999). Les données sont localisées sur une grille 50x50 km. Les 1200 espèces retenues (nombre d'occurrences > 20) incluent tous les ptéridophytes et une partie des spermatophytes avec toutes les familles des gymnospermes (Coniferales, Taxales et Gnetales), et une partie des angiospermes dicotylédones (Salicales, Myricales, Junglandales, Fagales, Urticales, Proceales, Santales, Aristologiales, Balanophorales, Polygonales, Centrospermae et Ranales) mais aucune monocotylédones.

La chorologie de chaque espèce a été renseignée grâce aux références suivantes : Bolòs et al. (1990), ECOFLORA (<http://www.york.ac.uk/res/ecoflora/cfm/ecofl/index.cfm>), Flora Europaea (Tutin *et al.* 1964-1993) et CATMINAT (<http://perso.wanadoo.fr/philippe.julve/catminat.htm>). 10 classes ont donc été définies : Alpine, Atlantique, Boréoalpines (boréosubalpine et Artico-alpine), eurosibérienne (médio-europe), Introduites, Méditerranéenne stricte, Méditerranéo-Atlantique (Méditerranéenne ayant une large part de distribution dans la région Atlantique), Méditerranéenne large (Espèce méditerranéenne dont la distribution excède largement la région Méditerranéenne), Littorale et plurirégionales (espèces dont l'aire de distribution occupe plusieurs régions biogéographiques).

Données de milieu : Les 7 variables bioclimatiques utilisées précédemment dans les autres analyses avec une résolution de 50x50 km. Le modèle de changements climatiques utilisé est le GCM HadCM3 avec le scénario A1 qui décrit un monde futur avec une croissance économique rapide et l'introduction de nouvelles technologies.

BIOMOD – projections futures : BIOMOD a été utilisé pour prédire les distributions actuelles et potentielles futures des espèces à 50x50 km. Les 4 types de modèles ont été évalués sur les données d'évaluation (30% des données originales) à l'aide de la courbe ROC. Pour chaque espèce, le modèle avec la plus forte AUC a été sélectionné et utilisé pour prédire la distribution actuelle et future (potentielle) à 50x50 km. Le pourcentage d'habitats perdus et potentiellement gagnés par chaque espèce a été estimé en calculant le nombre de pixels perdus dans le futur et le nombre de pixels potentiellement gagnés.

Propriété des niches des espèces : La position et l'amplitude de la niche des espèces ont été estimées à l'aide d'une analyse de co-inertie Outlying Mean Index (OMI : Doledec *et al.* 2000). Différemment des biens connues CCA et RDA (analyse canonique des correspondances et analyse de redondance), OMI ne fait aucune hypothèse sur la forme des courbes de réponse des espèces aux gradients de milieu et donne un poids égal aux sites avec beaucoup d'espèces et ceux pauvres en espèces. Le résultat de cette analyse décrit la position moyenne de l'espèce dans l'espace des variables (position de la niche), qui représente une mesure de la distance moyenne entre les conditions moyennes de la zone d'étude (ici l'Europe) et les conditions moyennes utilisées par l'espèce. La position de la niche a été estimée suivant les gradients (axes principaux) retenus. De même que la variabilité des conditions utilisées par les espèces sur chaque gradient a été calculée et assimilée à l'amplitude de niche des espèces. La prévalence (index de rareté géographique) des espèces a été calculée comme le ratio du nombre de présences de chaque espèce sur le nombre total de pixels (2289).

L'exposition géographique des pixels aux changements a été estimée pour chaque variable comme le coefficient de variation entre les conditions futures et actuelles.

Relation propriétés des niches – projections : Des modèles linéaires généralisés avec un processus de sélection de variables ont été utilisés pour sélectionner parmi les propriétés de la niche des espèces, celles qui expliquaient le mieux le pourcentage d'habitats perdus et gagnés par espèce.

composition des communautés existantes, ou bien encore le remplacement de communautés actuelles par des nouvelles. Plusieurs études paléocéologiques ont montré que, sur les temps géologiques, le fait que les espèces répondaient individuellement aux variations environnementales, avait provoqué des remplacements et modifications continus des communautés (Prentice *et al.* 1991; Graham 1992; Harrison & Prentice 2003; Webb 1992; Grayson 2000).

Cette dernière partie sera donc consacrée à deux thèmes : le premier portera sur l'estimation de la sensibilité des espèces européennes aux changements globaux. Le deuxième thème portera sur la variation spatiale de la perturbation du pool régional d'espèces par pixel de 10' suivant une série de scénarios de changements climatiques (mondial *versus* régional ; économique *versus* écologique). Dans ce thème sera aussi développé une analyse d'incertitude permettant d'évaluer l'influence des méthodes de modélisation *versus* l'influence des scénarios sur les projections de turnover du pool régional d'espèces par pixel.

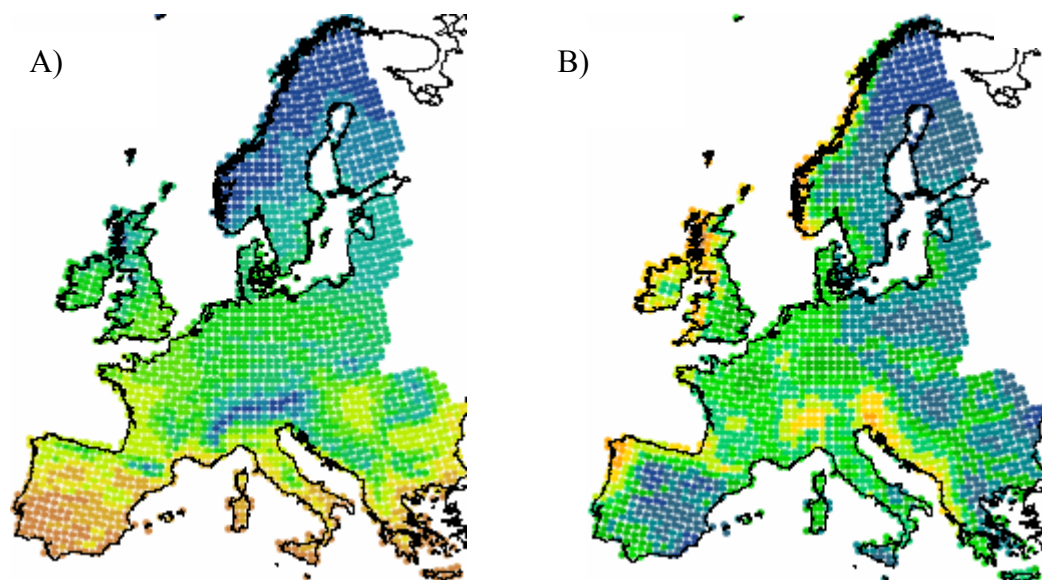
2 Sensibilité des espèces aux changements climatiques : Effets des propriétés des niches et de l'exposition géographique

Basé sur Annexe 3.1 :

Thuiller, W. Lavorel, S. & Araújo, M.B. Niche properties and geographic extent as predictors of species sensitivity to climate change. Soumis

En dépit du fait que des méta-analyses aient montré que les changements climatiques ont déjà affecté la biodiversité (Parmesan & Yohe 2003; Root *et al.* 2003), beaucoup d'efforts doivent encore être fait pour évaluer la sensibilité relative des espèces sur de grandes échelles spatiales. Les espèces très répandues, qui rencontrent une large gamme de conditions climatiques et édaphiques au travers de leur aire de répartition, sont supposées avoir une plus forte tolérance aux changements de milieu que les espèces à distributions plus restreintes (Brown 1995). En d'autres termes, les espèces avec des positions de niche proches des conditions de milieu moyennes et avec des amplitudes de niche assez larges devraient être moins sensibles que les espèces plus marginales et avec une relativement faible amplitude (Schoener 1989; Leibold 1995). Une espèce qui occupe différents types d'habitats (forte amplitude de niche) est génétiquement adaptée à toutes ces conditions environnementales, soit grâce à une forte diversité phénotypique entre populations, soit grâce à une forte plasticité. A l'inverse, une espèce typique d'un milieu particulier extrême aura développé des stratégies de reproduction et de survie très particulières la rendant inféodée à ce type de milieu (Thuiller *et*

Figure III.1 : Variation géographique des scores des deux axes de l'OMI en Europe. A) scores pour l'axe 1, B) scores pour l'axe 2. A) la gradation de rouge à bleu correspond à une baisse des températures ; A) la gradation de rouge à bleu correspond à une baisse des précipitations.



al. in press-c). Théoriquement, ces dernières espèces devraient donc être plus sensibles à des modifications de leur milieu d'ordre climatique ou d'utilisation des terres. Ceci devrait donc permettre de formuler des hypothèses simples sur la sensibilité des espèces aux changements climatiques en fonction de leurs seules caractéristiques de niches :

1 - Les espèces marginales (position de niche éloignée des conditions moyennes) devraient être plus sensibles à des variations de leur milieu que les espèces qui ont des positions de niches proches des conditions moyennes (Brown 1995)

2 - Les espèces avec de faibles amplitudes de niche (spécialistes) devraient être plus sensibles à des changements de milieu que des espèces généralistes (Swihart *et al.* 2003).

3 - Les espèces avec des aires de répartitions restreintes devraient être plus sensibles que des espèces à grandes aires de répartitions (Johnson 1998).

Néanmoins, toutes ces hypothèses dépendent de la distribution géographique des espèces (chorologie), de la direction des changements climatiques et de l'exposition des espèces aux changements supposés hétérogènes au travers des échelles spatiales (Walther *et al.* 2002). On a vu par ailleurs que les espèces de chorotypes différents ne sont pas supposées répondre aux mêmes facteurs du milieu (Thuiller *et al.* 2003b).

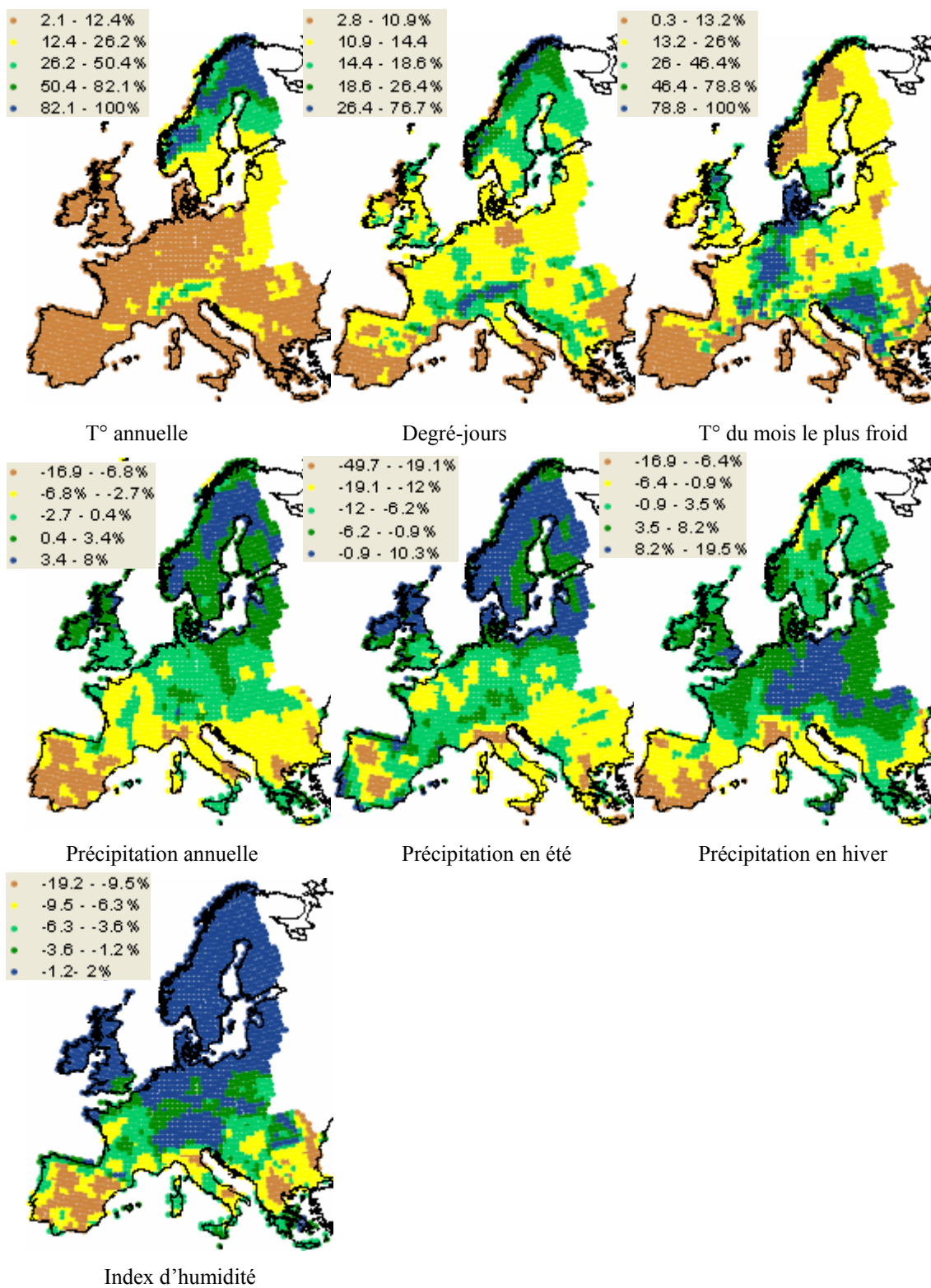
Nous avons donc mis en place une analyse dont le but était de tester ces différentes hypothèses sur les propriétés de la niche (position, amplitude et exposition aux changements climatiques) et de distribution géographique (chorologie et rareté géographique) comme indicateurs de sensibilité. Nous avons mis en relation les propriétés de la niche des espèces avec les simulations effectuées avec BIOMOD sous un scénario de changements climatiques (HadCM3 A1 mondial et économique avec doublement de la concentration en CO₂) (Encadré III.1).

L'OMI effectuée sur le couple de tableaux variables bioclimatiques x occurrences d'espèces montre qu'en Europe, la distribution des 1200 espèces de plantes sélectionnées s'explique suivant deux axes principaux, dont le premier très dominant (80%) peut être assimilé à un gradient de température et le second (13%) à un gradient de précipitation. Le gradient de température exprime un gradient nord-est/sud-ouest, tandis que le gradient de précipitation exprime un gradient de continentalité et d'altitude (Figure III.1).

2.1 Exposition des plantes européennes aux changements climatiques

A échelle continentale, l'exposition des espèces ou des zones géographiques est variable (Figure III.2). La sensibilité d'une espèce donnée dépendra fortement de sa distribution géographique et de ses propriétés écologiques, relativement à l'exposition aux changements

Figure III.2: Exposition géographique de l'Europe aux changements climatiques. Les pourcentages représentent les coefficients de variations entre le climat actuel et futur.



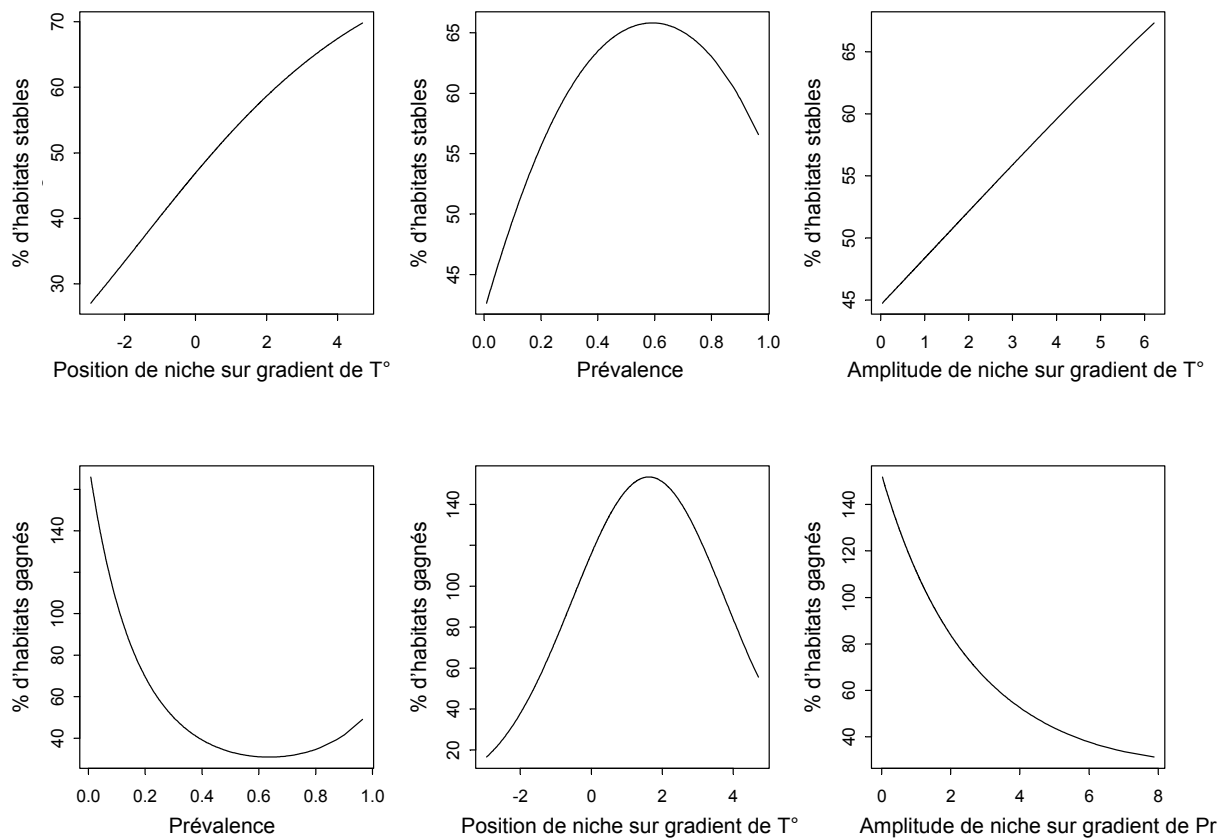
climatiques. Une espèce à l'intérieur d'une zone exposée aux changements sera forcément plus sensible que si elle était dans une zone non exposée. Similairement, dans une zone exposée, une espèce avec une forte amplitude de niche devrait conserver une plus grande partie de son habitat qu'une espèce avec une faible amplitude. A échelle européenne, un gradient de température très fort et un gradient de précipitation permettent de séparer les niches des espèces. Dans un contexte de réchauffement climatique, les espèces se situant préférentiellement dans les zones plus chaudes pourraient bénéficier de ces changements dans le futur, en conservant leurs habitats initiaux et/ou en s'étendant vers de nouveaux habitats potentiellement favorables. Inversement, les espèces plutôt inféodées à des climats humides et froids devraient être vulnérables, perdre une importante proportion d'habitats favorables et, si possible, migrer vers le nord de l'Europe ou monter en altitude.

2.2 Limitations méthodologiques

Dans les méthodes utilisées dans cette étude (BIOMOD et OMI), la niche écologique modélisée est supposée agir comme une contrainte stable sur la distribution géographique des espèces. L'hypothèse principale de ce type de modélisation est que les espèces modélisées sont à l'équilibre (ou quasi-équilibre) avec leur milieu (Prentice *et al.* 1992; Guisan & Zimmermann 2000). Dans cette étude comme dans toutes celles de cette thèse, j'ai émis l'hypothèse que les espèces étaient proches de l'équilibre compte tenu de l'échelle et de la résolution des données utilisées et des études paléocologiques qui soutiennent cette hypothèse (Prentice *et al.* 1991; Webb 1992). Ceci est particulièrement justifiable par le fait qu'à l'échelle considérée, l'impact de l'utilisation des terres et couverture du sol sur la distribution des espèces semble être masqué par les gradients macro-climatiques européens (Thuiller *et al.* in press-a).

Une autre limitation concerne la description complète des dimensions des niches écologiques des espèces : Premièrement, l'analyse présentée ici ne capture pas l'ensemble des limites bioclimatiques des espèces pour décrire avec fiabilité leur niche. Ceci est particulièrement le cas des espèces en marge de distribution telles que les espèces du sud-est de la Méditerranée. Comme nous ne possédons pas les données d'espèces et du climat d'Afrique du Nord (Maroc, Algérie, Tunisie), nous ne connaissons pas les limites de température supérieure (Tmax) de nombreuses espèces méditerranéennes présentes sur les deux continents. Deuxièmement, projeter la distribution de certaines espèces dans le futur au delà des conditions bioclimatiques utilisées pour calibrer le modèle provoque nécessairement une baisse de la fiabilité et des problèmes de sur- et sous-prédictions (Thuiller *et al.* in press-b). Troisièmement, le degré par

Figure III.3 : Courbes de réponse des % d'habitats stables et potentiellement gagnés en fonction des propriétés de niche sélectionnées par les GLM



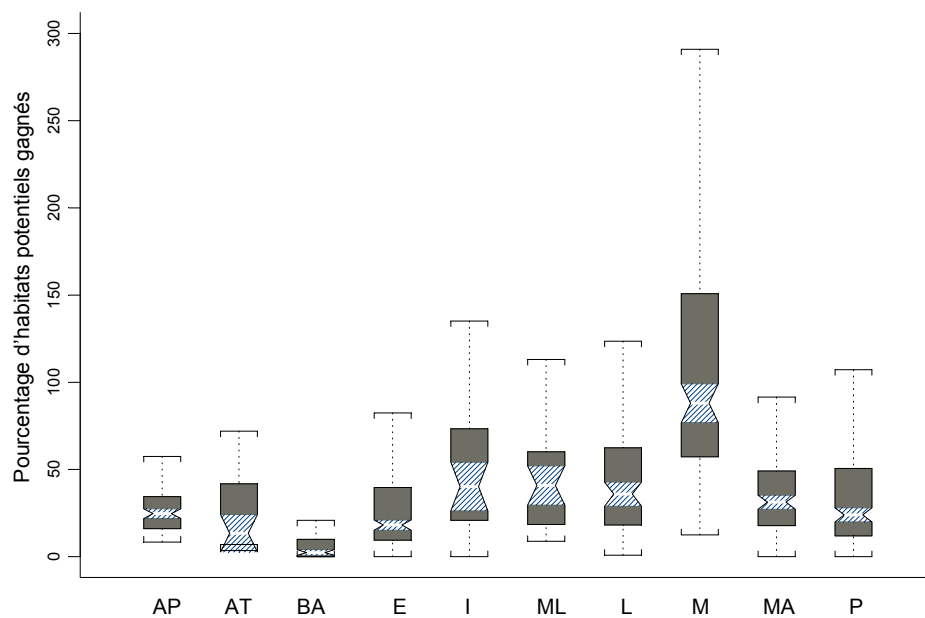
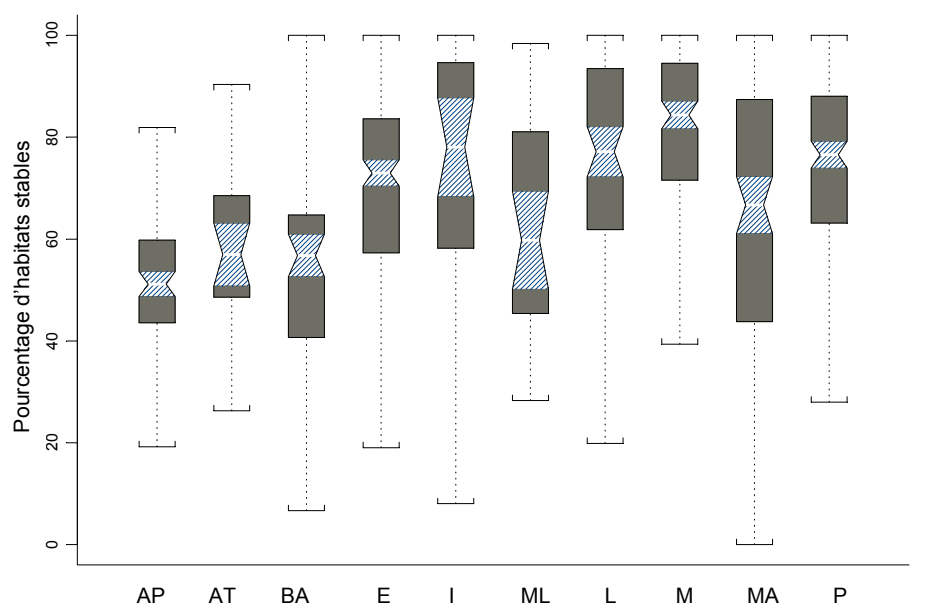
lequel les interactions interspécifiques peuvent limiter les distributions actuelles des espèces à grande échelle peut mener à sous-estimer la distribution actuelle potentielle et donc à perturber les projections des futurs habitats potentiels (Peterson 2003).

Les projections des habitats potentiels adéquats dépendent aussi fortement de la résolution de l'analyse. Ceci est principalement le cas des espèces alpines, où une résolution de 50x50 km ne peut prendre en compte l'hétérogénéité et la structure spatiale des montagnes (Araújo *et al.* soumis), faisant que les conclusions sur ces espèces doivent être émises avec précaution.

2.3 Est-ce que les espèces marginales sont les plus sensibles aux changements climatiques ?

L'analyse des relations entre la position de niche des espèces en Europe et leurs futurs habitats potentiels montre que les espèces marginales ne sont pas obligatoirement sensibles aux changements climatiques car leur sensibilité dépend aussi fortement de l'exposition, de l'amplitude de leur niche et de la taille de leur aire de répartition (Figure III.3-A). Des espèces marginales, comme certaines espèces méditerranéennes, sont même projetées être favorablement bénéficiaires des changements climatiques (Figure III.4-B). Ces résultats corroborent des analyses réalisées en Afrique du Sud dans lesquelles les espèces les plus vulnérables sont soit les espèces marginales avec des aires disjointes, soit des espèces localisées préférentiellement dans des zones arides de l'ouest, les plus exposées aux changements climatiques (Erasmus *et al.* 2002). La marginalité d'une espèce devrait donc toujours être mise en relation avec le gradient écologique considéré, susceptible de varier dans le futur. Si les changements climatiques ne modifient pas trop les relations qu'elles ont avec le milieu, et inversement créent de nouvelles conditions favorables, les espèces marginales peuvent potentiellement, sous réserve de dispersion et de colonisation, devenir des espèces en expansion. Mais inversement, si les changements climatiques vont à l'encontre de leurs préférences écologiques, ces espèces deviennent très vulnérables car elles utilisent des conditions de milieu marginales en Europe ce qui rend les possibilités de migration réduites. C'est le cas par exemple des espèces boréoalpines (Figure III.4) qui sont très marginales sur le gradient de température (coté froid). Ces espèces étant dans une zone très exposée avec des amplitudes de niche étroites, font que ce chorotype est particulièrement menacé de contraction vers le nord de la Scandinavie. Les mêmes conclusions peuvent ainsi être déduites des espèces alpines, excepté que la migration vers des altitudes supérieures peuvent leur permettre de résister aux changements climatiques (Gottfried *et al.* 1995; Peterson 2003). Ces fluctuations géographiques recourent les études paléoécologiques qui montrent que les espèces

Figure III.4 : Relations entre les chorotypes et (a) le pourcentage d'habitats (pixels) stables dans le futur et (b) le pourcentage d'habitats potentiellement gagnés dans le futur par les espèces. AP: Alpine; AT: Atlantique; BA: Boréoalpine; E: Eurosibérienne; I: Introduite; ML: Méditerranéenne large; L: Littorale; M: Méditerranéenne; MA: Méditerranéo-Atlantique; P: Plurirégionale. Les hachures correspondent aux intervalles de confiance à 95%.



boréoalpines qui ont occupé une grande partie de l'Europe durant la dernière ère glaciaire, sont en régression vers le nord depuis maintenant 10000 ans (Prentice 1986; Huntley *et al.* 1989). Ce phénomène de migration en '*va et vient*' des types de végétation nordiques tels que la Taïga a d'ailleurs été observé au cours de nombreux cycles glaciaires-interglaciaires sans pour autant que ces types de végétation disparaissent (Webb 1992).

2.4 Les espèces spécialistes sont-elles plus sensibles aux changements climatiques que les espèces généralistes ?

Il est généralement avancé que les espèces occupant de nombreux types d'habitats différents (généralistes), et qui donc rencontrent une large gamme de conditions climatiques et topographiques, ont une plus grande tolérance aux changements climatiques que les espèces d'écologie plus restreintes (spécialistes) (Brown 1995). Notre analyse sur le gradient de température (le plus susceptible d'être modifié) corrobore cette hypothèse. La relation linéaire entre le pourcentage d'habitats stables dans le futur et l'amplitude de niche sur ce gradient montre que les espèces occupant un vaste champ de conditions climatiques à travers leur aire de distribution devraient être les plus tolérantes à un changement climatique (Figure III.4). Dans un sens, ceci peut sembler évident, mais ces espèces ne sont pas forcément prédites gagner des habitats potentiels supplémentaires. Similairement à ce qui était attendu (Swihart *et al.* 2003), les espèces avec des amplitudes de niche réduites sont prédites perdre de grosses proportions d'habitats existant si elles se localisent préférentiellement dans des zones froides et exposées en Europe. Ceci est principalement le cas des espèces boréoalpines et quelques espèces eurosibériennes.

Tout comme les espèces du côté froid du gradient de température, les espèces avec une faible amplitude de niche du côté chaud sont aussi prédites perdre de grosse quantité d'habitats favorables sans pour autant en gagner. Ces espèces affectionnent des combinaisons de variables climatiques particulières qui ne devraient pas être plus communes dans le futur.

2.5 Est-ce que les espèces avec des aires de répartitions limitées sont plus sensibles aux changements climatiques que les espèces largement répandues ?

Les espèces rares occupent une place centrale en écologie et en conservation car elles sont le plus susceptibles à l'extinction (Pimm *et al.* 1988; Gaston 1994). Dans notre étude, nous n'avons pas abordé directement le risque d'extinction des espèces rares puisque nous

avons supprimé de nos analyses les plus rares d'entre elles. De plus, seule la rareté géographique, fonction de la taille de l'aire de répartition et non de l'abondance locale, a été considérée. Cependant, concernant les espèces avec des aires de répartition restreintes, les simulations mettent en évidence que les changements climatiques devraient contribuer à augmenter le risque d'extinction de ces espèces, résultant d'une forte régression de leurs habitats favorables, et ce, principalement pour les espèces se situant dans des régions exposées (climats froids et humides). Néanmoins, les espèces avec un faible taux d'occurrences sont supposées gagner des habitats potentiels favorables quand elles sont situées au milieu du gradient de température. Cette relation curvilinéaire entre la taille de l'aire de répartition et le pourcentage d'habitats potentiellement gagnés par les espèces (Figure III.3-B), ne tient cependant pas compte des distances géographiques et des capacités de dispersion des espèces considérées, pas plus que de la structure de la distribution géographique des espèces. Une espèce avec une aire de répartition limitée peut être avoir une distribution soit très clairsemée, soit très localisée. Dans le premier des cas, l'espèce a colonisé différents sites géographiquement éloignés, dont certains peuvent rester épargnés des changements climatiques. Cependant, cette espèce possède un faible nombre de populations à l'échelle régionale donc peu de sources de propagules pour pouvoir disperser et bouger spatialement avec son habitat. Les espèces les plus rares sont d'ailleurs souvent considérées comme ayant une capacité de dispersion limitée et n'étant pas de bons compétiteurs (Gaston 1994). Il est donc peu probable que ces espèces puissent atteindre les habitats potentiellement favorables modélisés. De plus, les espèces rares à échelle biogéographique, sont souvent localement rares en abondance et donc très influencées par les conditions locales et des changements tels que la fragmentation, la modification de l'utilisation des terres et les invasions par des espèces exotiques (Gaston 1994; Vitousek *et al.* 1997), ce qui rend plus difficile une migration vers de nouveaux habitats favorables. Dans le cas d'espèces à distribution agrégée, le risque d'extinction est aussi important si les changements climatiques provoquent une disparition ou une restriction des conditions de milieu adéquates. Cependant, un nombre de populations plus important que dans le cas de distributions clairsemées leur procure une banque de propagules plus importante et donc une capacité de dispersion plus importante.

3 Sensibilité des pools régionaux d'espèces aux changements climatiques : projections et incertitudes

On a vu que les modèles statiques permettent d'évaluer le risque que peuvent exercer les changements globaux sur certaines espèces en Europe. Si les analyses ont été réalisées sur un grand nombre d'espèces pour pouvoir trouver des patrons forts et extrapolables, il n'en reste pas moins que ces analyses ne portent que sur l'entité étudiée, ici l'espèce. Cependant on peut aussi évaluer les impacts de changements globaux sur le pool régional d'espèces dans une zone donnée (ex : pixel) et donc d'estimer indirectement l'impact des changements sur les communautés et les écosystèmes que constituent ces espèces. Si dans une région donnée, il y a un grand nombre d'occurrences d'espèces qui disparaissent par pixel, il est facile d'extrapoler que cette région va subir un bouleversement dans les habitats qui la composent. L'étude spatiale de la richesse spécifique par pixel, le pourcentage d'espèces perdues, stables ou gagnées par pixel, ainsi que le taux de turnover des espèces permettent d'évaluer l'impact des changements au niveau du pool régional d'espèces.

Un biais majeur de ces estimations est le nombre d'espèces étudiées. Il est évident que pour que cette mesure soit fiable à échelle grossière, il faudrait avoir des données pour toutes les espèces vivant dans un pixel donné, c'est à dire un échantillonnage exhaustif. Ceci est à ce jour impossible, et les seules bases de données existantes plus ou moins fiables à échelle européenne sont les quatre atlas présentés dans les chapitres précédents. Comme il s'agit de la meilleure représentation possible, on peut s'affranchir de ce biais en supposant que ces atlas, bien qu'incomplets, fournissent un échantillonnage représentatif, ou du moins hautement informatif, des communautés et des écosystèmes présents dans chaque pixel. Toute conclusion ou extrapolation future doit tenir compte de ce biais important.

De nombreuses études ont analysé les relations entre la richesse spécifique (RS) et des variables de milieu (ex : climat, fragmentation), pour tenter d'une part de comprendre si la RS est une propriété émergente des écosystèmes (Brown *et al.* 2001) ou si ce sont les écosystèmes qui influent sur la RS (Tilman & Pacala 1993; Tilman *et al.* 1997), d'autre part de prédire globalement les relations richesse-climat (Francis & Currie 2003), ou encore d'analyser les relations entre la richesse locale et le pool d'espèces régional (Koleff & Gaston 2002). Cependant, peu d'études ont tenté de prédire la RS dans un contexte de changements globaux (mais voir Saetersdal *et al.* 1998; Rutherford *et al.* 1999; Guisan & Theurillat 2000). La première difficulté vient du fait que la RS est un assemblage d'espèces non stable dans

l'espace et dans le temps. Etant donné que les études paléocéologiques (Webb 1981; Prentice *et al.* 1991, Harrison & Prentice 2003) et la théorie individualiste de Gleason (Gleason 1926) nous montrent que les espèces répondent de manière individuelle aux modifications de leur milieu, il paraît nécessaire de modéliser premièrement les distributions actuelles et futures des espèces pour estimer ensuite la RS actuelle et future par simple comptage du nombre d'espèces présentes par pixel (Austin 1998; Saetersdal *et al.* 1998, Guisan, 2000 #226). Cependant, il faut que les modèles réalisés au niveau spécifique soient fiables, car la somme des imperfections et des incertitudes peut devenir un bruit trop conséquent à estimer au niveau de la richesse ou autre estimateur basé sur la somme de modèles. On peut alors se demander quelle peut-être l'effet du choix des modèles utilisés pour l'estimation du turnover ou de la richesse spécifique par pixel dans le futur. Qu'elle est l'incertitude des projections liée aux modèles ou autres approches purement empiriques, comparée à l'incertitude ou variabilité d'utiliser différents scénarios de changements climatiques ?

Deux études seront donc présentées ici : la première portera sur l'effet des différentes méthodes de modélisation *vs* les différents scénarios de changements climatiques, sur les projections futures de turnover d'espèces dans des pixels de 10'. La seconde étude sera basée sur les projections du pourcentage d'espèces gagnées, perdues et du turnover des espèces dans le futur, suivant différents scénarios climatiques après la prise en compte de l'incertitude méthodologique.

3.1 Incertitudes des projections dans le futur

Basé sur Annexe 3.2 et 3.3:

Thuiller, W., Araújo, M. B., Brotons, L., Lavorel, S, Midgley, G. & Williams, P.H. (En préparation).

Uncertainties in species range shifts projections under global climate change scenarios.

Thuiller, W. (Soumis) Patterns and uncertainties of species' range shifts under global change.

Des études récentes, utilisant des modèles basés sur les niches comme présentés en première partie, ont exposé des conclusions assez alarmistes sur les effets des changements globaux sur la faune et la flore (Bakkenes *et al.* 2002; Pearson *et al.* 2002; Peterson *et al.* 2002b). Par exemple, le pourcentage de turnover d'espèces prédit au Mexique pourrait atteindre 40%, avec une disparition de la faune comprise entre 27 et 75% par pixel en 2050 (Peterson *et al.* 2002a). De même, en utilisant l'Atlas Florae Europaeae (20% de la flore européenne), Bakkenes *et al.* (2002) projettent que 32% des espèces de plantes modélisées en Europe devraient être en déclin en 2050.

Encadré III.2

Données espèces : Les quatre atlas présentés dans les autres paragraphes. Après avoir supprimé les espèces dont le nombre d'occurrences était inférieur à 20, 2013 espèces ont été modélisées (383 oiseaux nicheurs, 162 mammifères, 118 reptiles/amphibiens, 1305 plantes).

Données de milieu : Les sept variables bioclimatiques utilisées dans Encadré BIOMOD.

BIOMOD : pour chaque espèce, ANN, CTA, GLM et GAM ont été calibrés sur 70% des données originales et évalués sur les 30% restant. Les quatre modèles ont ensuite été utilisés pour projeter les distributions actuelles et les futurs habitats potentiels des espèces à 10° suivant 5 scénarios d'émissions (HadCM3 A1-A2-B1-B2 et CSIRO2 A2). Pour chaque espèce, les 20 projections dans le futur ont été transformées en présence-absence en utilisant soit la courbe ROC soit la statistique du Kappa. 40 projections ont donc été établies pour chacune des 2013 espèces.

Projections : Chaque espèce est prédite gagnée ou perdre des pixels adéquats. Pour évaluer les patrons géographiques de pertes et gains en Europe, la somme d'espèces perdues (P) et gagnées (G) par pixel est calculée. Après avoir déterminé la richesse spécifique par pixel (RS), le pourcentage de turnover peut être estimé suivant la méthode : $100(P+G)/(RS+G)$

Incertitude : L'incertitude peut-être mesurée comme l'écart à la moyenne générale du turnover. Cette mesure de variabilité ne permet pas cependant d'étudier quel composant influe le plus sur cette incertitude. Nous avons donc analysé l'incertitude des projections à l'aide d'une analyse de consensus. Cette analyse est basée sur une analyse principale des correspondances (PCA) sur les 40 vecteurs de projections du turnover. Si les 40 projections étaient identiques, le pourcentage de variance expliquée par le premier axe serait de 100%. A l'inverse, si les 40 projections étaient aléatoires, le pourcentage de variance expliquée serait égale à 1 divisé par le nombre de variables : $1/40 = 2,5\%$.

La variation expliquée par le 1^{er} axe résume la similarité des projections. Pour analyser les sources d'erreurs, les axes orthogonaux au premier axe sont examinés. Comme les composants Modèles x Méthodes Seuils x Scénarios sont plus ou moins emboîtés, il est difficile avec une ANOVA de mesurer les sources d'incertitudes. Pour cela nous avons utilisé des arbres de classifications qui permettent de prendre en compte les interactions multiples entre composants et qui ne sont pas perturbées par les problèmes d'emboîtement étant non paramétriques. Les arbres permettent donc d'avoir une vue hiérarchique de l'influence de chacun des trois composants de l'analyse à l'incertitude.

Bien que les modèles basés sur les niches aient été intensivement comparés sur leur capacité à prédire la distribution actuelle de nombreux taxons à différentes échelles spatiales (Guisan *et al.* 1999; Moisen & Frescino 2002; Thuiller *et al.* 2003a), leur fiabilité dans le futur reste évidemment plus difficile à évaluer. Dans la seconde partie de ce manuscrit, j'ai montré qu'une gamme de modèles avec des puissances prédictives similaires sous des conditions actuelles pouvait prédire des distributions futures très différentes (Thuiller 2003). De tels résultats ont d'importantes implications quand les projections pour des espèces individuelles sont ajoutées pour dériver des taux de turnover ou d'extinction par unité de surface (ex : pixel) comme utilisés par Bakkenes *et al.* (2002) et Peterson *et al.* (2002a). En effet, si pour une espèce, les projections dans le futur de deux modèles (ex : GAM et ANN) diffèrent notablement, il y a de fortes chances que ces différences soient amplifiées en utilisant un plus grand nombre d'espèces. Il y a donc un besoin urgent d'évaluer l'impact de la variabilité entre modèles pour pouvoir dériver des conclusions sur l'impact des changements globaux sur la diversité spécifique. Ce besoin est de plus augmenté quand les projections dans le futur sont réalisées en utilisant une série de scénarios de changements climatiques, comme recommandé dans le dernier rapport du GIEC (Watson & *et.* 2001). La question est donc :

Comment la variabilité d'utiliser une série de scénarios de changements climatiques, représentant différentes réponses de la société, est-elle confondue par la variabilité résultante de l'utilisation de différentes techniques de modélisation ?

Afin d'y répondre, nous avons construit une analyse pour prédire le pourcentage de turnover d'espèces en Europe, basée sur les quatre atlas européens de plantes, mammifères, reptiles & amphibiens et oiseaux. Pour analyser différentes sources de variabilité, nous avons utilisé BIOMOD pour prédire la distribution des 2013 espèces à l'aide de modèles ANN, CTA, GLM et GAM. A ce stade, il y a donc 4 prédictions pour chaque espèce. Ensuite, pour analyser la variabilité des projections suivant les scénarios employés, nous avons utilisé 2 modèles GCM (HadCM3 et CSIRO2) ajustés aux quatre scénarios d'émissions proposés par le GIEC (A1, A2, B1, B2). Le modèle CSIRO2 n'étant disponible que pour le scénario A2, il y a donc 5 projections futures différentes. Pour chaque espèce, selon le modèle utilisé et le scénario, il y a donc 20 projections dans le futur. Enfin, pour étudier les distributions futures des espèces et donc calculer le pourcentage de turnover ou la richesse spécifique, il faut transformer les probabilités d'occurrences générées par les modèles en données présence-absence. Nous avons utilisé la courbe ROC et la statistique du Kappa pour dériver le seuil

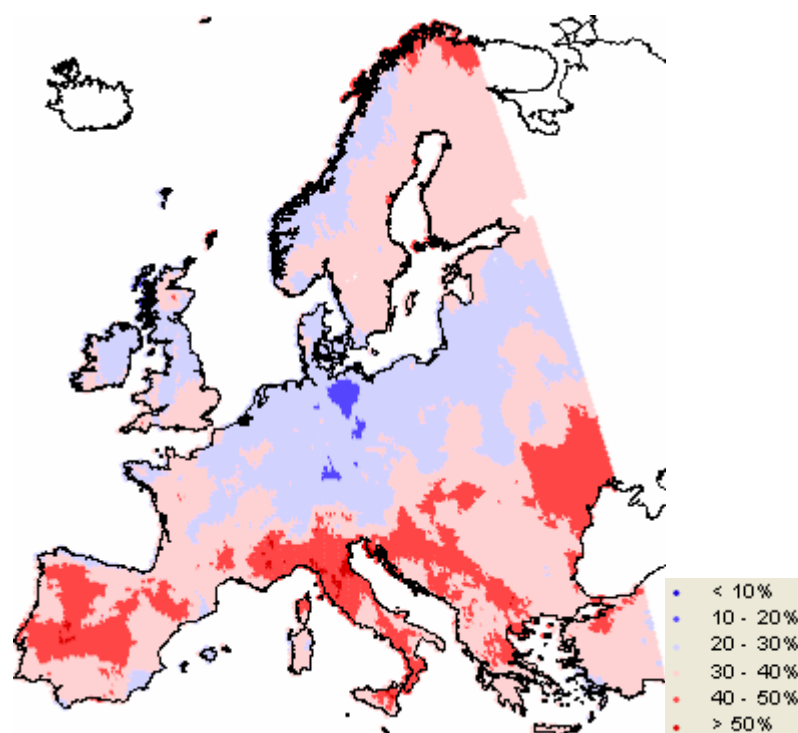
Tableau III.1 : Minimum (Min), 1er quartile, médiane, moyenne, 2nd quartile et maximum du turnover moyen sur l'Europe.

	Min	1 ^{er} Quartile	Médiane	Moyenne	2 nd quartile	Max
Turnover	15,8	28,5	33,5	33,3	37,8	54,5

Tableau III.2 : Pourcentage d'explication des premiers axes de la PCA

	Axe 1	Axe 2	Axe 3	Axe 4
Variance	5,16	2,12	1,58	1,40
Proportion variance expliquée	0,665	0,112	0,063	0,057
Proportion cumulée x 100	66,5	77,7	84,0	90,0

Figure III.5 : Pourcentage de turnover moyen d'espèces à travers les 40 projections



nécessaire à la transformation en données d'occurrences. Au final, pour chaque espèce, il y a donc 4 (modèles) \times 5 (scénarios) \times 2 (méthodes seuils) = 40 projections en 2050.

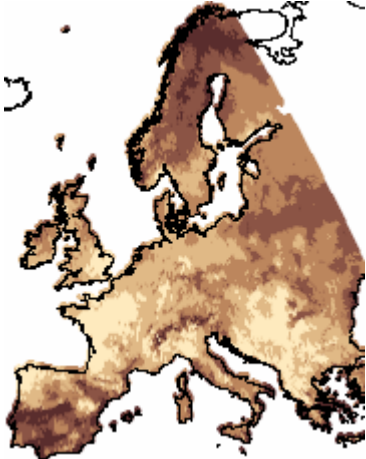
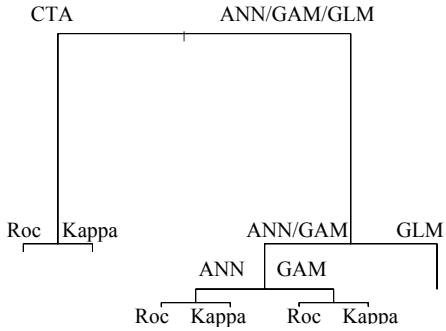
Ensuite pour chaque triplet (modèle - scénario - méthode seuil), nous avons estimé le pourcentage d'espèces potentiellement disparues et gagnées par pixel ainsi que le pourcentage de turnover spécifique (Encadré III.2).

Le turnover moyen par pixel à travers les 40 projections est très variable (Tableau III.1) avec un maximum atteignant les 55%. Cette variabilité se retrouve d'une manière spatiale où certaines régions devraient être très peu perturbées par les changements climatiques (ex : Nord de l'Allemagne), tandis que d'autres devraient subir un très fort bouleversement dans le pool régional d'espèces (Centre-ouest de la Péninsule Ibérique, sud-est de la France, nord de l'Italie et ex-Yougoslavie) (Figure III.5). Certaines parties isolées sont même prédites subir une modification de plus de 50% de la faune et de la flore actuelle (Frontière Espagne-Portugal, quelques zones des Alpes Maritimes françaises ou de l'Italie). Cependant, cette moyenne est peu informative compte tenu de la variabilité liée aux différents composants. On peut toutefois remarquer que les zones de fort turnover sont respectivement la région Boréo-Alpine, très exposé à un réchauffement et la région méditerranéenne, exposée à une forte aridification (Figure III.2).

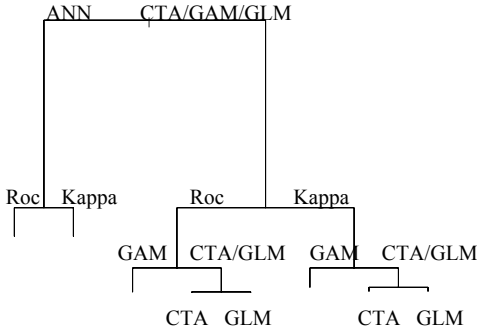
L'analyse de la PCA dite de consensus, permet de comprendre la similarité entre projections. Les quatre premiers axes de l'analyse expliquent 90% de la variabilité entre projections (Tableau III.2). Le premier axe expliquant 66,5% de la variabilité, on peut dire que les projections sont similaires dans 66,5% des cas. Dans le meilleur des cas, le pourcentage restant d'incertitude (33,5%) devrait résulter de l'utilisation de différents scénarios de changements climatiques car ils intègrent de la variabilité certaine. Les axes suivants fournissent l'explication concernant cette variabilité. Les arbres de classification, réalisés pour mettre en évidence les composants qui expliquent les axes orthogonaux au premier axe de consensus, montrent que (Figure III.6) : Premièrement les scénarios de changements climatiques ne sont pas sélectionnés comme variables explicatives sur les axes 2 et 3, où les modèles expliquent la plus grosse part de variations, suivis par les méthodes seuils. Même si le but ici n'est pas d'explicitier les différences entre modèles, on peut toutefois noter que les différences majeures se situent sur les extrêmes en Europe (sud de l'Espagne, sud de la Grèce et nord de la Scandinavie)(Figure III.6 - Carte des scores des axes). Les modèles se distinguent certainement compte tenu de leur habilité à modéliser les espèces en marge de distributions ou les bornes extrêmes de distributions des espèces. Ce n'est que sur l'axe 4, qui n'explique que

Figure III.6 : Arbres de classification sur les deuxième, troisième et quatrième axes de l'ACP consensus et projections géographiques de ces axes.

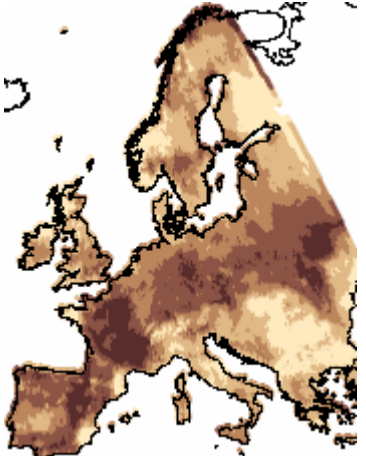
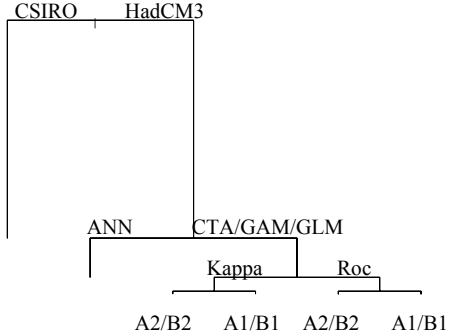
Axe 2



Axe 3



Axe 4



5% de la variabilité des projections, que les scénarios apparaissent comme variable explicative. La différence majeure se situe entre les GCM utilisés (HadCM3 et CSIRO), ce qui était attendu et ensuite suivant les scénarios d'émissions, séparant les scénarios dits 'globaux' des scénarios à tendance régionale. La variabilité spatiale entre ces scénarios et GCM s'exprime nettement sur un axe central européen de manière assez homogène (Sud de l'Espagne - Pologne).

Les résultats de cette analyse d'incertitude sont d'une part inquiétants et d'autre part révélateurs. Ils sont inquiétants parce qu'ils montrent qu'une étude utilisant un seul type de modèle peut s'avérer biaisée et surtout que les conclusions ne sont pas facilement généralisables compte tenu de la part de variabilité due aux méthodes empiriques (ex : Bakkenes *et al.* 2002). D'une manière plus générale, cela signifie que les incertitudes les plus grandes pour les projections futures sont issues des outils et non de la variabilité liée au fait qu'on étudie des systèmes naturels. Ces résultats sont révélateurs dans le sens où ils confortent l'utilité d'une méthode comme BIOMOD pour tester différents modèles pour avoir une amplitude de projections. Cependant, pour limiter le nombre de projections, il convient de trouver une méthode qui permettrait de sélectionner un modèle ou une méthode par espèce qui serait le plus en accord avec les autres dans le futur, et non pas sélectionner un modèle sur des données actuelles (Kappa ou courbe ROC) sans savoir comment vont répondre les modèles quand ils sont appliqués à des scénarios futurs. Une alternative serait donc de pouvoir sélectionner un modèle consensus et conservatif par espèce, qui ne serait pas forcément le meilleur mais le plus en accord avec les autres.

3.2 Sensibilité des pools régionaux d'espèces par pixel aux changements globaux

Basé sur Annexes 3.3. et 3.4.

Thuiller, W. (Soumis) Patterns and uncertainties of species' range shifts under global change

Thuiller, W. Araújo, M.B., Lavorel, S. (In preparation) Impact of global climate change on distributions of European taxa: projections under a set of IPCC storylines.

On a vu que l'incertitude des projections sur les 50 prochaines années était principalement liée aux méthodes empiriques utilisées et que cette incertitude cachait la variabilité d'utiliser différents scénarios de changements climatiques. Ceci est très dommageable car le GIEC a développé une série de scénarios d'émission censés englober différentes combinaisons de développements socio-économiques et de changements

Encadré III.3

Données espèces : même que l'analyse d'incertitude du paragraphe précédent. 2013 espèces dont 1305 plantes (10% flore européenne), 383 oiseaux nicheurs (80% avifaune européenne), 162 mammifères (80% faune mammifère européenne) et 118 reptiles & amphibiens (80%). Les données sont digitalisées sur une grille UTM 50x50.

Variables de milieu : même que l'analyse d'incertitude ou que dans l'encadré 2.2. 7 variables bioclimatiques. Pour réduire l'incertitude liée aux modèles de changements climatiques, nous avons utilisé le plus courant et le plus développé HadCM3, avec les quatre scénarios d'émissions A1, A2, B1 et B2. Les explications et détails sur ces scénarios sont exposés en Encadré II.1

Analyses : Comme dans l'analyse sur l'incertitude, BIOMOD a été utilisé avec les 4 modèles ANN, CTA, GAM et GLM pour prédire la distribution actuelle des espèces sélectionnées et projeter leurs distributions futures pour chaque scénario. Ensuite pour chaque scénario, les projections ont été transformées de probabilité d'occurrences en données présence-absence grâce aux deux méthodes courbe ROC et statistique du Kappa. Pour chaque scénario, il y a donc 8 projections par espèces.

Le modèle le plus conservatif et résumant l'information concomitante des autres a été choisi comme le plus corrélé au premier axe de l'ACP réalisée sur les 8 projections.

Projections : pour évaluer l'impact des changements climatiques au niveau des habitats (pixels), le nombre d'espèces disparues (conditions climatiques favorables actuellement mais plus dans le futur), et gagnées (conditions climatiques non favorables actuellement pour l'espèce donnée mais qui le seront dans le futur) a été calculé pour chaque pixel et exprimé en pourcentage par rapport à la richesse spécifique actuelle par pixel. Le pourcentage d'espèces perdues par pixel est égal à 1 - le pourcentage d'espèces stables par pixel.

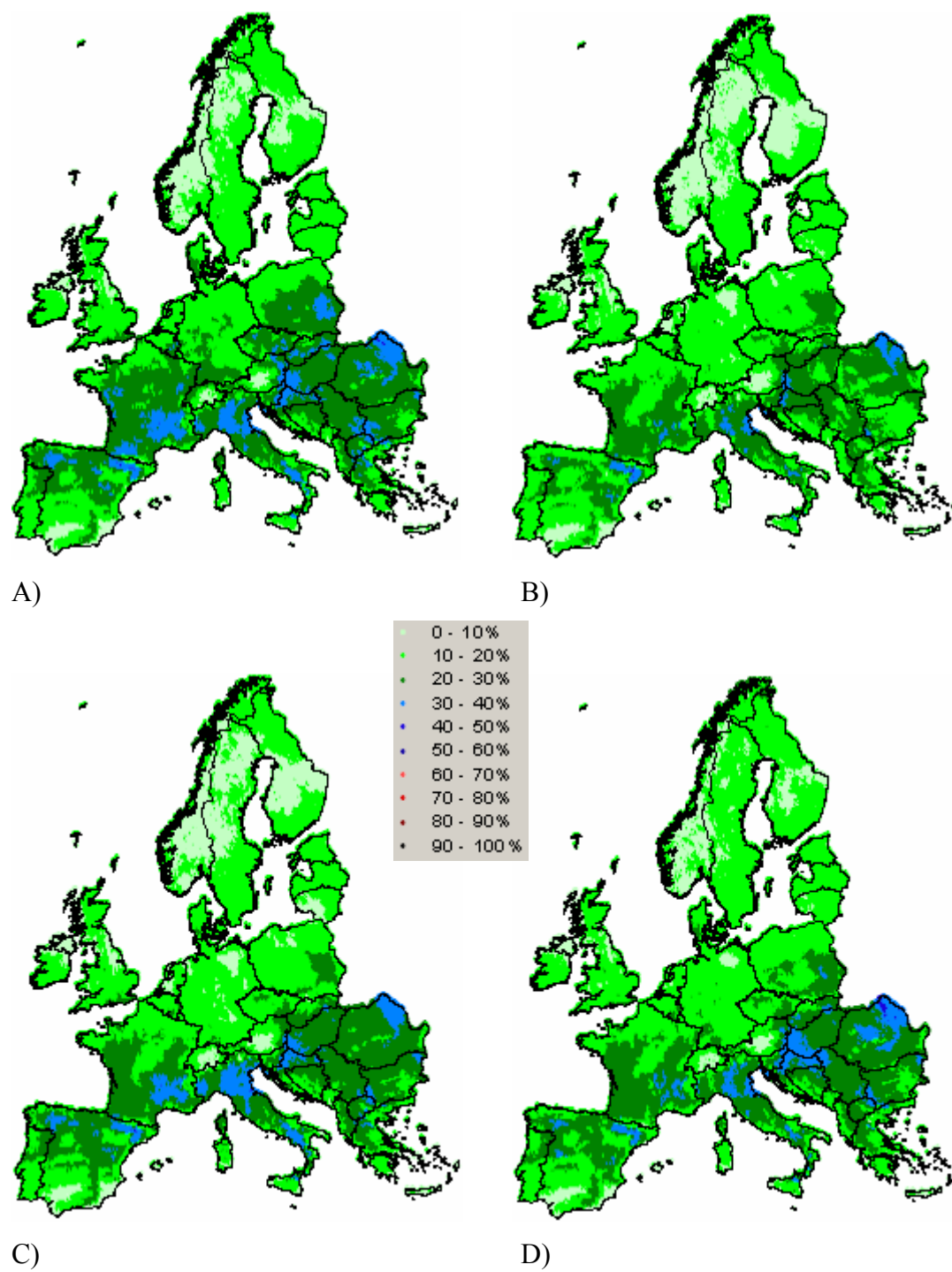
démographiques. Il est donc utile d'évaluer l'impact des changements climatiques suivant ces scénarios pour comprendre et prédire quel peut être le scénario le plus menaçant pour la biodiversité, et quel peut être le moins préjudiciable. Il est évident que ces extrapolations restent hypothétiques mais elles permettent d'ouvrir un nouveau champ de prédiction en fonction des développements régionaux et économiques des régions concernées.

Pour conclure cette thèse, il semble donc important d'évaluer l'impact des changements climatiques sur une partie de la biodiversité européenne en fonction de différents scénarios, ceci d'une manière géopolitique (par pays) et d'une manière biogéographique (par région biogéographique). Cependant, pour cela, il faut pouvoir prendre en compte l'incertitude explicitée dans le paragraphe précédent. Pour pouvoir s'affranchir de cette incertitude, il faut pouvoir choisir, par espèce, un modèle qui représente une sorte de compromis entre les autres, comme une sorte de moyenne. Cependant, cette moyenne est difficilement extrapolable et il est difficile de pouvoir étudier écologiquement les réponses des espèces aux variables du milieu. De plus, une moyenne doit être accompagnée d'un coefficient de variation pour pouvoir apporter de l'information pertinente. Afin d'explicitier ces résultats aux observateurs ou décideurs, il vaut mieux pouvoir trouver une méthode plus simple et surtout minimisant les cartes et sorties de modèles. L'utilisation de la moyenne, plus le coefficient de variation n'est pas donc retenue, même si elle reste une méthode intéressante. Pour pouvoir déterminer quel peut être le modèle le plus en accord avec les autres, c'est à dire résumant l'information la plus pertinente de tous les modèles, j'ai choisi d'utiliser l'analyse de consensus présentée dans le paragraphe précédent (Lebreton, J.-D. comm. pers.). Cette méthode permet de déterminer le modèle le plus en accord avec les autres, comme celui qui est le plus corrélé avec le premier axe de l'ACP. Comme le but est d'analyser les projections suivant les scénarios de changements climatiques, l'analyse de consensus a été réalisée au niveau 'modèles x méthodes seuils' de l'analyse d'incertitude. C'est à dire que seulement 8 projections par espèce sont réalisées par scénarios (4 modèles x 2 méthodes seuils), et une seule sera gardée comme étant la plus conservative (Encadré III.3).

3.2.1 Limitations biologiques et méthodologiques

Les limitations de cette analyse sont similaires à celles déjà émises précédemment. Cependant, parmi elles, deux limitations apparaissent ici prépondérantes. Premièrement, le fait de ne pas capturer la niche bioclimatique totale des espèces. En Europe, un certain nombre d'espèces sont en limite d'aire de distribution. C'est principalement le cas des espèces d'Europe de l'est, des régions irano-turaniennes et pontiques. Compte tenu que ces espèces

Figure III.7 : Taux potentiel d'extinction par pixel pour les 2013 espèces à travers l'Europe. A) scénario A1; B) scénario A2; C) scénario B1 et D) scénario B2.



sont en limite d'aire de distribution, il est probable que les modèles aient des difficultés à faire une bonne approximation de leur niche bioclimatique réelle et à projeter leurs futurs habitats potentiels avec précision (Thuiller *et al.* in press-b). Même si ces espèces sont minoritaires, cela peut avoir des répercussions importantes quand les distributions sont ajoutées pour évaluer la richesse spécifique géographique et l'impact des changements climatiques. Les conclusions sur ces régions doivent être émises avec précaution et en connaissance de cause. Deuxièmement, les régions alpines ou possédant un relief très hétérogène ne sont pas bien représentées avec des résolutions de 50x50 km ou de 10'. Pour avoir un aperçu plus fiable de ce qui se passe dans ces régions, des analyses plus fines doivent être utilisées (Gottfried *et al.* 1995 ; Guisan *et al.* 1998). Dans notre étude, la possibilité des espèces à monter en altitude ou à trouver des refuges n'est pas prise en compte. Pour une même augmentation de température (ex : 1°C), une espèce de montagne devra migrer très localement (1° de perte tous les 100m) pour retrouver une température adéquate tandis qu'une espèce de plaine devra certainement migrer beaucoup loin pour subsister (Peterson 2003). Comme précédemment, compte tenu de l'échelle et de la résolution utilisées, toute conclusion sur ces zones doit être sujette à précaution.

3.2.2 Taux d'extinction locale par pixel en Europe

Le taux d'extinction potentiel par pixel pour toutes les espèces montre des différences spatiales entre scénarios assez marquées (Figure III.7). Les fortes corrélations entre la richesse spécifique actuelle et les taux d'extinction locale montrent que les zones avec une forte diversité seront celles le plus probablement affectées par les changements climatiques (Test de corrélation de Spearman, ρ moyen sur les 4 scénarios = 0.791). Les scénarios A1 et B1 pourraient fortement affecter la richesse spécifique actuelle en provoquant les plus forts taux d'extinction locale par pixel. Inversement, les scénarios A2 et B2 qui sont supposés représenter un monde porté sur la conservation des identités régionales sont moins critiques sur les taux d'extinction locale par pixel même si dans certaines zones, ce taux peut atteindre 40% (Figure III.7). Bien que très intéressants, ces taux d'extinction restent purement informatifs car les espèces utilisées pour cette analyse ne représentent qu'un échantillon de la diversité totale en Europe. Les mêmes estimations au niveau des frontières géopolitiques montrent des dissimilitudes importantes, tant au niveau des scénarios, qu'au niveau des estimations d'extinction (Tableau III.3). Par exemple, la Hongrie, la Roumanie, la Macédoine ou la Croatie sont projetée perdre entre 20 et 30% de la faune et flore étudiées dans cette analyse, tandis que des pays du nord de l'Europe continentale comme la Norvège, la Hollande

Tableau III.3 : Pourcentage d'espèces disparues (Ext) et potentiellement gagnées (Col) selon le scénario dans les pays européens. Les chiffres en gras représentent les maximums par pays et ceux en gras italique les minimums.

Pays	A1		A2		B1		B2	
	Ext	Col	Ext	Col	Ext	Col	Ext	Col
Albanie	19.8	25.6	15.4	27.1	20.8	21.7	18.9	24.2
Allemagne	17.7	23.1	13.5	16.5	13.3	16.3	14.9	15.7
Autriche	19.2	22.4	14.8	19.5	16.0	18.1	17.2	19.1
Belgique	17.9	24.2	14.4	19.7	14.6	18.7	15.9	19.2
Bosnie H.	23.3	36.7	20.4	39.2	22.2	31.5	23.3	35.9
Bulgarie	24.1	27.8	19.6	29.1	23.7	25.6	23.2	28.9
Croatie	26.9	40.1	23.7	42.6	26.3	33.1	26.7	40.7
Danemark	18.2	23.0	15.7	21.3	15.3	23.3	16.4	21.2
Espagne	19.9	26.9	18.4	23.0	20.4	24.1	19.7	24.0
Estonie	17.5	38.2	15.6	32.9	15.2	38.6	15.6	35.3
Finlande	13.0	40.6	11.7	36.8	11.7	39.6	13.4	42.5
France	24.2	33.8	20.0	27.2	21.7	28.0	21.4	28.4
Grèce	17.2	21.8	14.5	21.5	17.1	19.0	17.2	20.8
Hongrie	28.0	46.5	23.7	47.7	27.2	37.2	28.1	49
Irlande	12.9	20.5	11.3	19.7	12.2	19.6	12.7	21.9
Italie	23.1	33.3	20.6	34.1	23.8	30.0	22.0	32.3
Lettonie	16.0	33.7	14.2	29.8	13.9	34.8	14.4	30.2
Lituanie	13.4	29.9	11.3	26.8	10.8	32.6	12.0	26.6
Luxembourg	22.8	27.3	18.2	22.8	18.0	21.4	19.3	22.7
Macédoine	27.1	28.0	21.3	28.5	25.5	25.2	25.0	28.3
Monténégro	18.4	37.6	15.1	42.2	17.8	33.1	18.6	36.9
Norvège	11.0	29.9	10.8	29.1	11.7	29.1	12.7	33.0
Pays-Bas	12.6	21.8	10.7	18.8	10.9	19.1	12.2	18.4
Pologne	20.6	27.2	16.5	21.4	15.5	21.1	17.9	20.9
Portugal	17.6	24.6	16.3	19.8	18.0	21.2	17.3	20.1
Roumanie	26.8	37.8	22.9	37.5	25.8	32.6	27.5	38.5
Royaume U.	13.5	23.5	12.2	21.5	12.9	21.7	13.1	22.9
Serbie	22.1	39.8	25.2	32.8	26.3	39.9	25.7	35.4
Slovaquie	25.7	35.4	21.5	32.4	21.5	27.0	23.6	32.2
Slovénie	23.5	28.2	20.7	30.6	20.9	22.1	23.1	28.1
Suède	12.1	35.4	11.0	33.5	11.2	32.7	12.5	37.8
Suisse	16.6	21.5	13.1	21.6	13.0	19.0	13.9	19.7
Tchéquie	22.9	32.1	17.8	24.7	17.3	22.9	18.6	24.4
Moyenne	19.6	30.2	16.7	28.2	18.0	26.7	18.6	28.3
Ecart- type	4.9	6.9	4.3	7.9	5.2	7.0	4.9	8.2

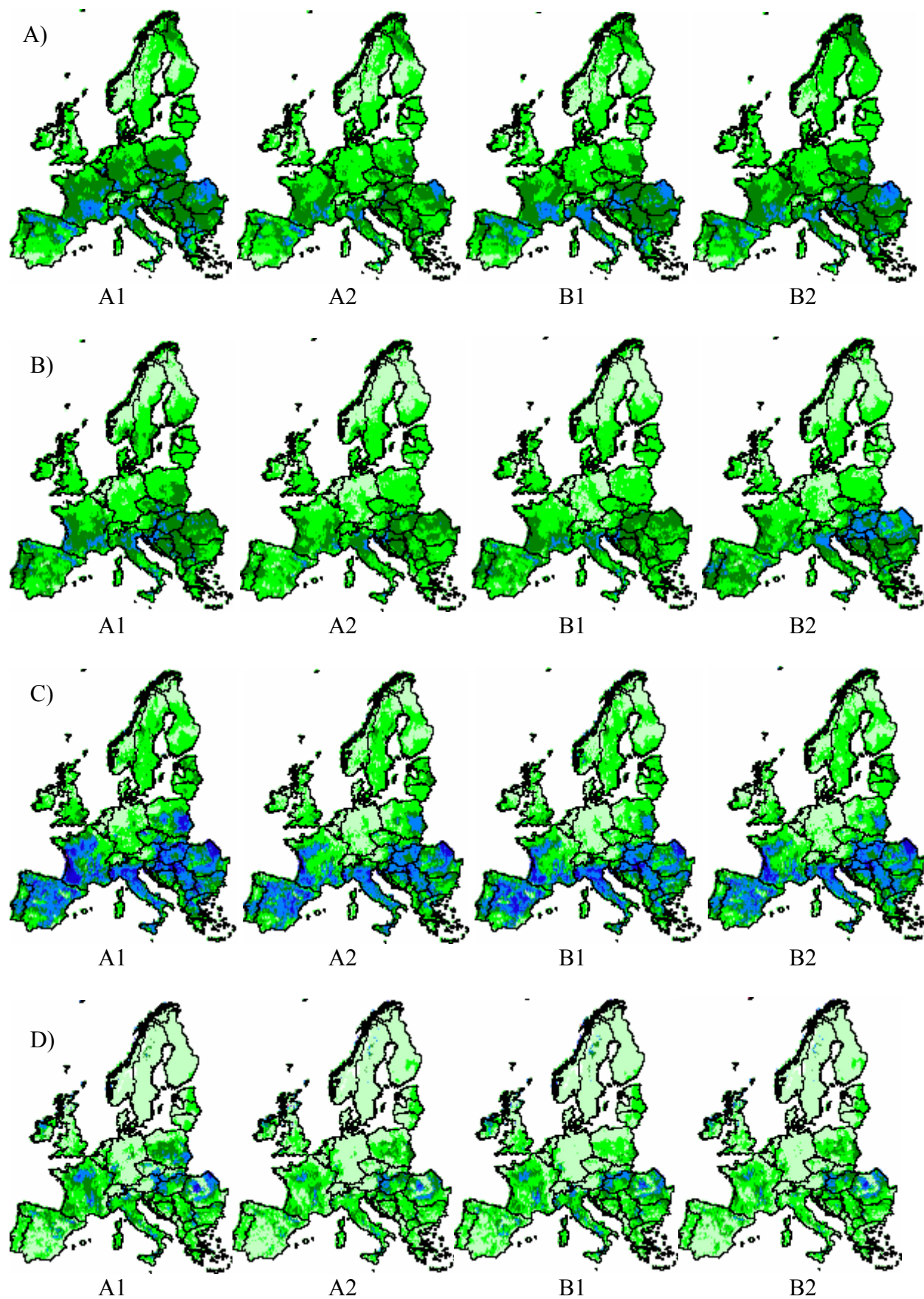
ou l'Allemagne sont projetés perdre entre 10 et 20% des espèces étudiées. Dans le même temps, la Tchéquie est projetée perdre en moyenne 23% des espèces étudiées par pixel sous le scénario A1 contre 17% sous le scénario A2.

Même s'il est difficile de générer une hiérarchie des scénarios en rapport à l'impact qu'ils ont sur les espèces, le scénario A1 est celui qui semble le plus préjudiciable. La différence entre les autres scénarios peut sembler trop faible et trop hétérogène pour tirer des conclusions valides seulement sur les espèces étudiées. Cependant, comme les espèces ne se répartissent pas d'une manière géopolitique, il paraît plus pertinent d'analyser les patrons d'extinctions au niveau des régions biogéographiques européennes (Tableau III.4). Curieusement, le fait que le scénario A1 provoque les taux d'extinction locale les plus forts n'apparaît plus aussi évident. Inversement, dans les régions alpines, arctiques et pannoniennes, c'est sous le scénario B2 que le taux d'extinction est le plus fort. Le scénario A1 est le plus préjudiciable dans les régions atlantiques, boréales, continentales et méditerranéennes (les plus dominantes en Europe). En ce qui concerne la région steppique, ce sont les scénarios B1 et B2 qui provoquent les plus forts taux d'extinction locale par pixel. Il y a donc une redéfinition de l'influence des scénarios. Même si le scénario A1 peut sembler le plus préjudiciable pour la biodiversité (Encadré II.1), il ne l'est que sur les régions dominantes en Europe mais pas sur les régions extrêmes. L'influence des scénarios au niveau de la diversité totale étudiée évolue donc d'une manière biogéographique.

Le taux d'extinction par groupes taxonomiques montre des patrons notables (Figure III.8). Premièrement, même si cela peut paraître trivial, la réponse des espèces des divers groupes taxonomiques est différente, même si la région méditerranéenne reste toujours la plus touchée. Le taux d'extinction chez les plantes et oiseaux (Figure III.8-A-B) est assez similaire quantitativement et spatialement. La France, l'Italie et la région pannonienne sont les zones critiques d'extinction (Tableau III.4), avec des taux compris entre 20 et 30%. Curieusement, cette similitude, seulement basée sur des relations modélisées entre la niche des espèces et les variables de milieu utilisées, devraient être renforcée compte tenu du fait que la végétation est l'habitat principal pour les oiseaux, que ce soit pour les sites de pondaisons ou de recherche de nourriture (Chamberlain *et al.* 1999; Benton *et al.* 2003).

La distribution spatiale et biogéographique (Figure III.8-C ; Tableau III.4) des extinctions chez les mammifères diffère considérablement des plantes et oiseaux. Selon tous les scénarios, la majeure partie de l'Espagne, la France, l'Italie et toute la région Pannonienne devraient avoir de fortes pertes d'espèces par pixel, comprises entre 30 et 50%. Ces résultats corroborent l'analyse sur les relations niche-sensibilité, où les espèces eurosibériennes

Figure III.8 : Taux potentiel d'extinction par pixel pour les quatre groupes taxonomiques considérés. A) Plantes ; B) Oiseaux nicheurs ; C) Mammifères et D) Reptiles et amphibiens. A1, A2, B1, et B2 correspondent respectivement aux scénarios utilisés (voir Encadré II.1)



devraient subir une réduction de leur aire de répartition par le sud. L'Allemagne, la Suisse, l'Autriche et la Belgique devraient subir peu de pertes par pixel et donc avoir des distributions assez stables. Pour finir, les taux d'extinctions chez les reptiles et amphibiens (Figure III.8-D; Tableau III.4) sont beaucoup plus faibles que les autres groupes taxonomiques (entre 0 et 40%). Seules quelques zones devraient être plus fortement touchées comme la région parisienne, la Tchéquie et la Hongrie (environ 30% de pertes par pixel).

Toutefois il est à noter que les observations suggèrent que les mammifères, reptiles et amphibiens sont les taxons qui montrent les plus forts taux d'extinction et de restriction d'aire ces dernières années. Ce clivage avec les projections peut s'expliquer par le fait que ces espèces sont soumises à de forts stress locaux par l'urbanisation croissante, le développement du trafic routier, des pressions anthropiques et de l'utilisation des terres. Ces paramètres ne sont pas pris en compte dans nos modèles, ce qui explique ces différences notables.

3.2.3 Taux de colonisation locale par pixel en Europe

Les phénomènes observés étant voisins de ceux décrits en ce qui concerne les taux d'extinctions, j'explicitai et discuterai plus rapidement les résultats. Les faibles corrélations entre la richesse spécifique actuelle et les taux d'extinction locale montrent les colonisations potentielles ne devraient pas être nécessairement dans des zones faiblement ou fortement diversifiées. (Test de corrélation de Spearman, ρ moyen sur les 4 scénarios = 0.283)

Selon les quatre scénarios utilisés, deux régions (Scandinavie et une zone centrale en Europe) se distinguent où les taux de colonisation sont prédits être les plus importants (Figure III.8). Tout comme pour les taux d'extinction, la Hongrie semble être le pays pour lequel le taux de colonisation potentielle est le plus fort (40% du pool d'espèces actuelles)(Tableau III.3). Le scénario A1 est celui sous lequel les taux de colonisation locale potentielle est le plus important à travers l'Europe, alors que le scénario B1 est celui qui produit le moins de possibilité de migration (Tableau III.3).

Chez les plantes, on peut voir que les régions qui devraient avoir le plus fort taux de colonisation sont les régions boréoalpines et alpines (Figure III.10-A). Cette prédiction rejoint d'une part les observations de pollen pendant le réchauffement climatique durant l'Holocène, où les espèces boréales ou/et eurosibériennes sont remontées vers la Scandinavie (ex : *Picea*, Huntley *et al.* 1989), et d'autre part les observations récentes résumées par Parmesan & Yohe (2003). En ce qui concerne les oiseaux, les deux principales zones où les colonisations pourraient être les plus importantes sont la région extrême nord de la Scandinavie et le sud d'est de la France, Nord de l'Italie et les Balkans (Figure III.10-B). Les prédictions de

Tableau III.4 : Pourcentage d'espèces disparues (Ext) et potentiellement gagnées (Col) par taxon et selon le scénario dans les régions biogéographiques européennes. Mamm = Mammifères. x représente une valeur inexistante. Il n'y a pas d'espèces de reptiles ou d'amphibiens actuellement ni projetées dans le futur dans les régions concernées.

Région	Taxon	A1		A2		B1		B2	
		Ext	Col	Ext	Col	Ext	Col	Ext	Col
Alpine	Total	15	31.5	13.3	30	14.4	29.3	15.8	32.4
	Plantes	16	33.8	14.1	32	15.5	32	17.4	33.4
	Oiseaux	14.2	23.9	12.8	23.1	13.4	21.3	13.3	24.5
	Mamm	15.5	37.4	14.3	36.1	15.1	35.1	14.7	38.2
	Reptiles	x	x	x	x	x	x	x	x
Arctique	Total	14.3	39.2	14	35.5	15.9	38.4	17.3	38.8
	Plantes	14.8	44.3	15.1	40.2	16.4	44.7	18.4	41.4
	Oiseaux	13.5	22.7	11.5	19.4	14.7	21.5	16.1	18.1
	Mamm	15.8	62.4	17	61.1	19.1	56.3	18.2	65.1
	Reptiles	x	x	x	x	x	x	x	x
Atlantique	Total	16.9	25.8	14.5	21.8	15.6	22.6	15.8	23.2
	Plantes	17.1	27.9	14.7	23	16	24	16.8	24.2
	Oiseaux	16.1	18.1	13.8	16.6	14.8	17	13.3	16.5
	Mamm	19.9	23.9	16.2	21.8	16.7	22.2	15.9	23.1
	Reptiles	x	x	x	x	x	x	x	x
Boréale	Total	13.9	28.6	12.4	35.5	12.4	37.3	13.7	39.4
	Plantes	14.5	53.2	13.3	49.2	13.2	52.9	16.2	51
	Oiseaux	13.7	20	11.8	18.9	12	19.1	10.6	19.3
	Mamm	14.9	38.2	13.3	34.6	12.8	36	13.5	37.5
	Reptiles	x	x	x	x	x	x	x	x
Continentale	Total	23.4	30.5	19.4	26.3	19.8	24.8	21.8	26.6
	Plantes	25.1	36.9	20.7	31.7	21.4	29.7	23.3	31.1
	Oiseaux	19.6	18	16.9	15.8	16.5	15.1	18.6	13.2
	Mamm	26	22.9	21.6	20.8	21.1	19.9	22.2	21.2
	Reptiles	18.4	55.6	14.5	45	14.1	44.6	15.4	47.7
Méditerranéenne	Total	19	25	17.3	23.2	19.5	21.9	19.1	23.1
	Plantes	25.1	36.9	20.7	31.7	21.4	29.7	23.3	31.1
	Oiseaux	19.6	22.8	17.1	21.3	18.7	20.2	20.3	15.7
	Mamm	25.7	23.9	24.5	23.1	25.5	20	23.7	24.6
	Reptiles	14.3	33.8	12.3	30.4	13.9	31	12.1	33.3
Pannonienne	Total	28	44.7	23.6	45	26.9	36	28.1	46.4
	Plantes	27.3	47.1	22.5	49	27.6	37.2	26.8	47.6
	Oiseaux	27.1	29.6	23.6	27.5	22.9	25.3	30.7	29.1
	Mamm	37.8	40.8	33.5	43	34.6	31.7	35.7	44.4
	Reptiles	23.8	119	22.9	104	24.2	98.8	22.7	82.1
Steppique	Total	29	41.3	27.8	45.5	28.6	36.7	30	45
	Plantes	29.5	45.1	28	52.9	29.8	39.5	29.6	48.9
	Oiseaux	26.7	27.4	26.1	26.1	24.8	25.5	30.9	26.8
	Mamm	37.6	37.7	36.1	44.1	34.1	33.6	36.9	42.5
	Reptiles	24.4	86.3	22.7	82.1	26.4	78.2	21.6	80.5

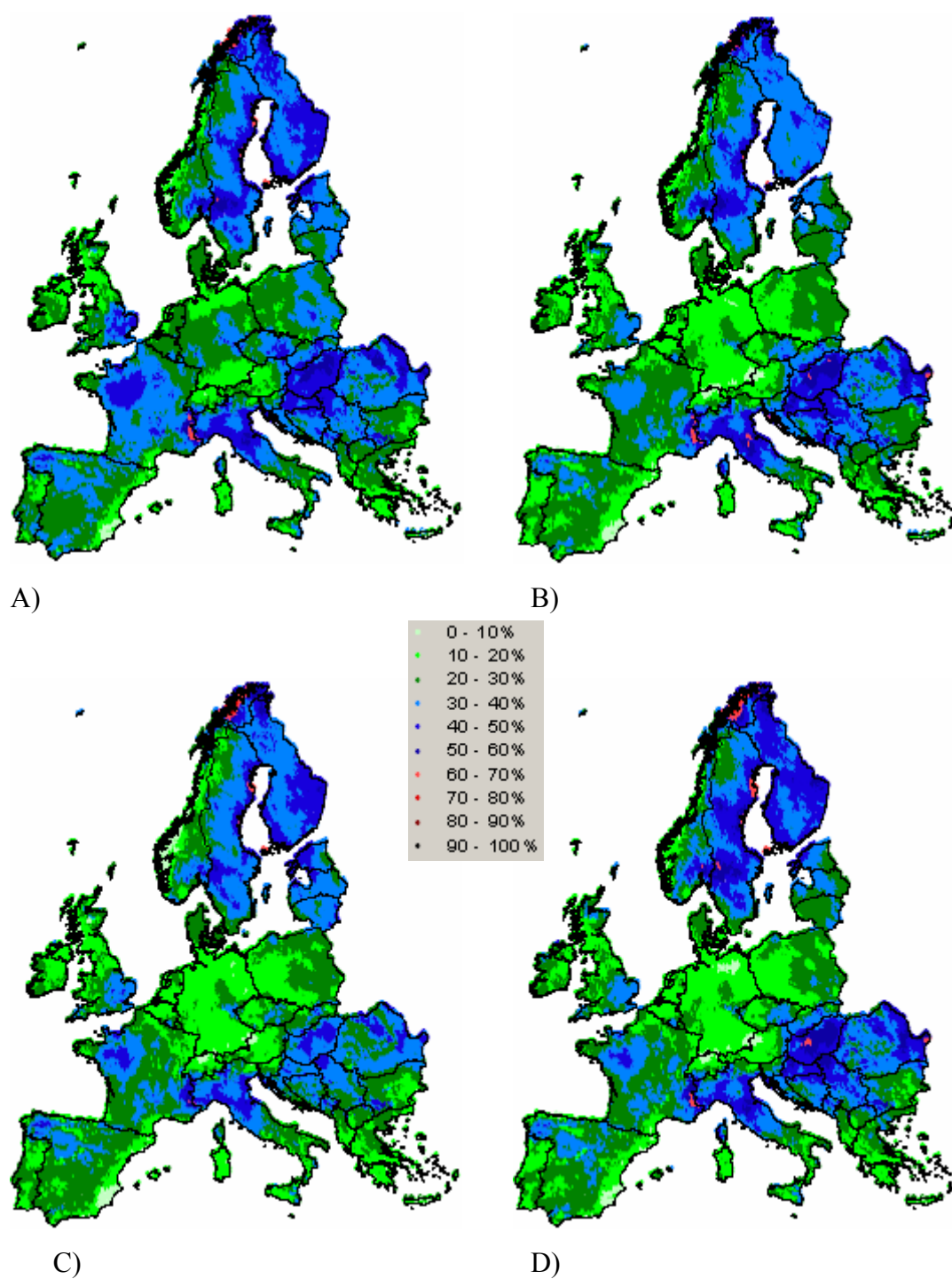
colonisations locales pour les mammifères sont proches de celle des oiseaux, mais avec des taux plus fort dans toute la Scandinavie (Figure III.10-C). Enfin, le réchauffement climatique montre que les reptiles et amphibiens pourraient avoir des taux de colonisation importants dans toute l'Europe (Figure III.10-D).

3.2.4 Complications écologiques

Les taux d'extinction et de colonisation locale sont calculés ici d'une manière statique, simplement déduit de l'affinité des espèces à des combinaisons de variables de milieu. Il convient toutefois de rappeler que diverses composantes écologiques influent ces taux. Par exemple, le taux d'extinction locale calculé par pixel ne prend en compte ni la plasticité phénotypique, ni les temps de réponse à la migration des organismes. Il est ainsi généralement admis que les arbres ont une plasticité plus importante que les herbacées, et peuvent donc s'adapter, au moins temporairement, à de nouvelles conditions de milieu. Des études paléoécologiques ont aussi montré que le temps de réponse moyen des arbres à un réchauffement similaire à celui prédit dans le futur, était compris entre 100 et 1000 ans (Prentice *et al.* 1991). Cependant, le réchauffement passé qui a produit de tels taux s'est déroulé sur un laps de temps d'environ 3000 ans comparé aux 100 ans actuellement prédits. Ces temps de réponse passés ne peuvent donc pas être facilement extrapolés au présent compte tenu de l'exceptionnelle rapidité des changements actuels et futurs (Webb 1992). Cette rapidité implique que les processus de sélection naturelle ne devraient pas permettre une sélection des génotypes les mieux adaptés par les modifications du milieu, du moins pour les espèces longévives. Les taux d'extinction et de colonisation locales dans un pixel dépendent donc de plusieurs paramètres écologiques : dispersion, plasticité, réponse à la sélection. Dans l'état actuel des choses, il est malheureusement impossible de savoir si les taux d'extinction prédits dans cette analyse sont fiables, compte tenu de l'imprévisibilité de ces mécanismes écologiques.

Il en va de même pour les taux de colonisation par pixel. Pour qu'une espèce colonise des nouveaux habitats, il faut tout d'abord que ces habitats soient favorables et qu'elle ait une capacité de dispersion suffisante. Les capacités de dispersion des espèces sont très difficiles à mesurer convenablement, car elles dépendent de facteurs intrinsèques à l'espèce (ex : mobilité, poids et nombres des propagules) et extrinsèques (ex : fragmentation, agent de dispersion, facilitation). Un vaste débat est en cours sur les taux de migration des arbres mesurés entre 100 et 1000 mètres par an à la fin du Pléistocène (Cain *et al.* 1998 ; Clark 1998). Plusieurs études ont montré que ces taux extrêmes pourraient être dus aux phénomènes rares de dispersion à

Figure III.9 : Taux potentiel de colonisation par pixel pour les 2013 espèces à travers l'Europe. A) scénario A1; B) scénario A2; C) scénario B1 et D) scénario B2.



longue distance (Cain *et al.* 1998 ; Clark 1998). Ces dispersions composites par saut (ou *dispersion leptokurtique*) engendrent rapidement, bien au delà du front de migration, des petits îlots de populations, qui ont un rôle d'accélérateur de la progression et du maintien de la variabilité génétique. Ces dispersions longues distantes restent cependant très difficilement quantifiables, les rendant difficiles à prédire. De même, il a été observé que certains mammifères avaient des distributions actuelles distantes de plus de 1500 km de celles qu'ils avaient juste après la dernière glaciation (Graham 1992). Même s'il est évident que les espèces répondent rapidement et sur de longues distances aux variations de milieu telles que les changements climatiques, il est aussi difficile de prédire ces colonisations en raison de la rapidité des changements actuels et prédits pour le futur.

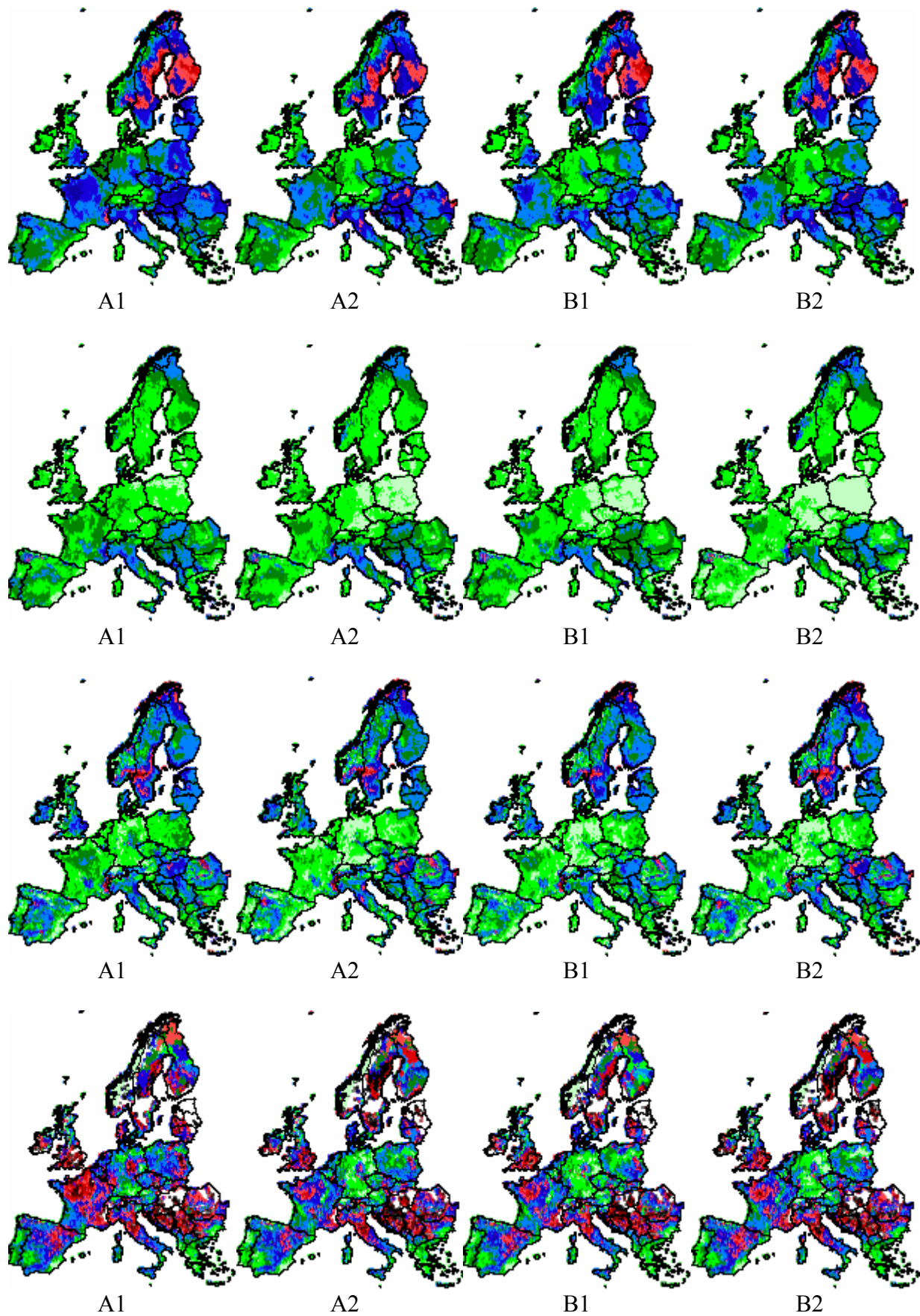
Dans un pixel où les taux de colonisations et d'extinctions sont forts, les communautés risquent d'être fortement perturbées même si ces taux ne sont pas effectivement réalisés. En effet, si le taux d'extinction locale est fort, de nombreuses espèces vont se retrouver en dehors des conditions climatiques qui leur sont adéquates. Ces espèces peuvent migrer (en ce cas on se rapproche du taux d'extinction locale prédit) avec un certain laps de temps, soit elles peuvent subsister grâce à une certaine plasticité. Dans ce cas là, il est probable que les rapports de dominances à l'intérieur des communautés soient modifiés. Par exemple, admettons que certaines espèces très compétitrices dans leur habitat ne peuvent pas suivre leur habitat dans l'espace. Certaines espèces, meilleurs colonisatrices et d'habitude exclues de cet habitat vont s'y implanter parce que l'espèce dominante ne sera plus la pour les exclure. Il se pourrait donc que la niche réalisée des espèces change en conséquence d'un remaniement des communautés et donc des interactions interspécifiques.

3.2.5 Patrons biogéographiques de sensibilité

Selon tous les scénarios, il apparaît une zone centrale en Europe, qui devrait être le théâtre de forts bouleversements spécifiques. Cette région qui part du nord de l'Espagne, vers le centre de la France, le nord de l'Italie, l'Autriche, la Hongrie et l'ex-Yougoslavie, est prédite subir de forts taux d'extinction et de colonisation locales par pixel. Inversement, le nord de l'Europe continentale (Allemagne, Belgique, Hollande ou Lettonie), ainsi que les zones à climat très méditerranéen sont prédites subir peu de changements (Figure III.11).

Cette zone centrale, où un fort turnover d'espèces devrait avoir lieu, représente les écotones entre les régions biogéographiques européennes et les types chorologiques qui les composent. En effet, dans le premier paragraphe, on a vu que les espèces eurosibériennes devraient subir des régressions de leur aire de répartition par le sud, tandis que certaines

Figure III.10 : Taux potentiel de colonisation par pixel pour les quatre groupes taxonomiques considérés. A) Plantes ; B) Oiseaux nicheurs ; C) Mammifères et D) Reptiles et amphibiens. A1, A2, B1, et B2 correspondent respectivement aux scénarios utilisés (voir Encadré II.1)



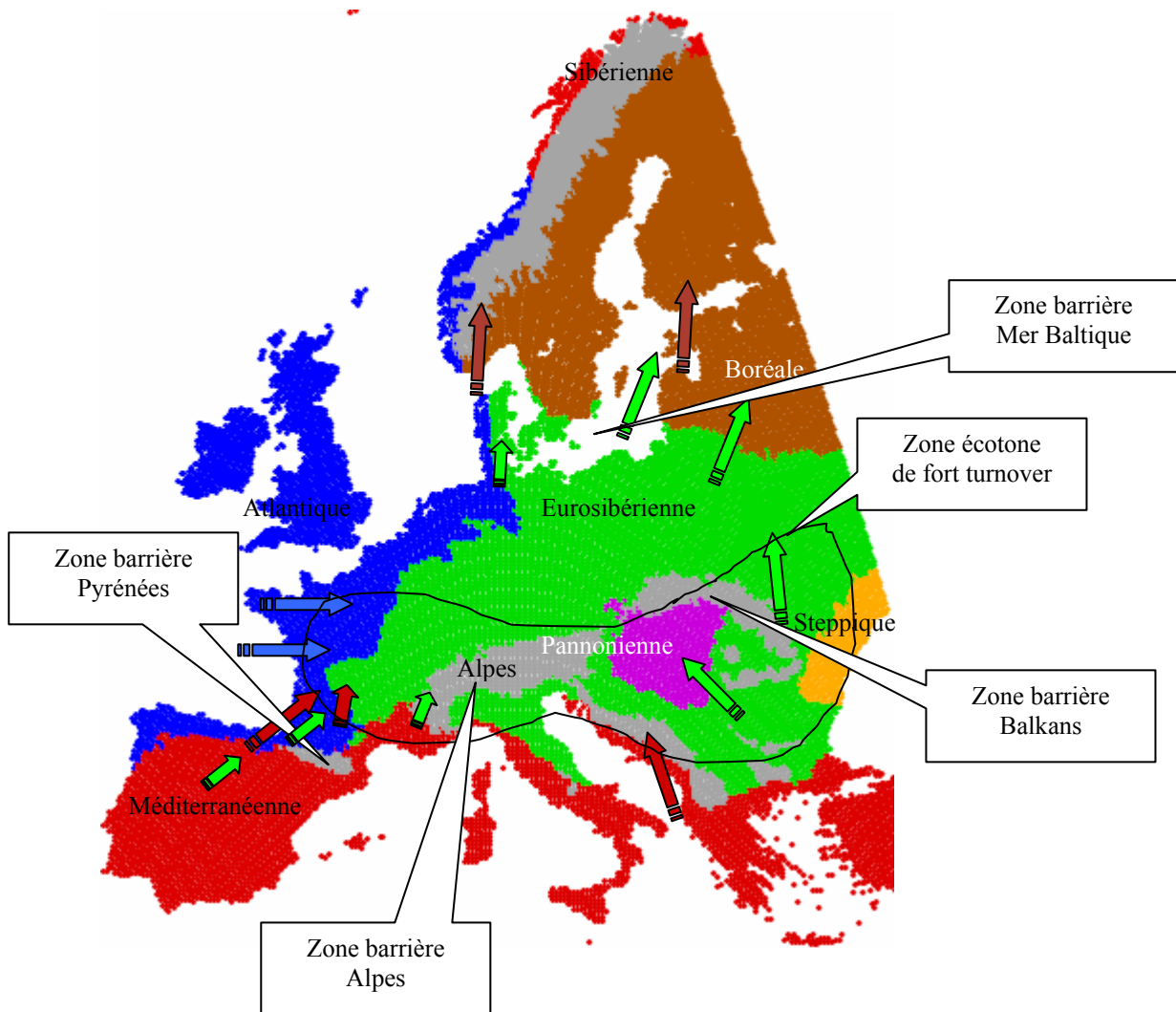
espèces méditerranéennes ne devraient pas trop être perturbées. La zone écotone sud de fort changement est donc une zone de contraction de distribution des espèces eurosibériennes et d'expansion des espèces méditerranéennes. La zone écotone ouest (parallèle à l'Atlantique) est une zone de contraction des espèces eurosibériennes et d'expansion des espèces Atlantiques (Figure III.11). De la même manière, les espèces eurosibériennes devraient s'étendre vers le sud de la Scandinavie, ce qui explique un fort taux de colonisation dans cette région, conjointement à un taux important d'extinction dû à la contraction des espèces boréoalpines. On peut supposer que ces zones sont hautement sensibles aux changements, même compte tenu de l'imprédictibilité de la réponse des espèces. Il a de fortes incertitudes d'une part sur les taux de pertes réels *versus* la plasticité de certaines espèces ou la survie de certaines dans des micro-habitats favorables (futurs espèces reliques), d'autre part sur les taux réels de colonisation qui pourront être continus sur 50 ans, et donc du rôle des interactions biotiques qui pourraient avoir lieu au sein des nouveaux assemblages formés.

On peut aussi se poser la question du rôle de l'utilisation des terres et de la fragmentation du paysage à échelle locale sur les taux de colonisations. En effet, si des taux de migrations de certains arbres compris entre 100 et 1000 mètres par ans ont pu être observés durant le Pléistocène, on peut émettre des doutes sur une telle vitesse dans des paysages européens sans cesse remodelés par l'homme. Toutes ces incertitudes rendent les prédictions difficiles à faire hormis le fait que ces zones devraient être perturbées et certainement subir un fort brassage spécifique et génétique.

De la même manière, on peut se demander le rôle de barrières comme les chaînes de montagnes ou la mer dans ce processus. Les Alpes, les Pyrénées et les Balkans sont des montagnes très hétérogènes qui proposent des conditions climatiques proches de celle du nord de l'Europe. Il est donc très probable que des espèces eurosibériennes tendent à coloniser des habitats particuliers des Alpes, qui créent ainsi des zones refuges pour ces espèces (Blondel & Aranson 1999). A l'inverse, ces barrières montagneuses, qui bloquent les migrations, pourraient créer un déficit de diversité juste au nord de ces massifs. Durant les dernières cycles glaciaires - interglaciaires glaciations, les Pyrénées et les Alpes ont joué un rôle prédominant dans la survie des taxons de certaines familles, en servant de refuges et permettant à beaucoup d'espèces de persister en procurant des conditions écologiques temporairement absentes dans leur milieu d'origines (Taberlet *et al.* 1998; Blondel & Aranson 1999).

La Figure III.10 montre que pour les plantes, les taux de colonisation locale pourraient être importants en Scandinavie (Sud de la Suède et Finlande). Outre les changements de distributions probables des espèces Boréoalpines vers le nord, ces taux reflètent aussi des

Figure III.11 : Région biogéographique actuelle et mouvement schématisé hypothétique des chorotypes associés européens.



habitats nouvellement favorables pour des espèces eurosibériennes. Comme dans le cas des barrières montagneuses, les espèces eurosibériennes devront naturellement traverser la Mer Baltique pour coloniser ces régions ou passer par la Finlande. Un autre point intéressant est la situation des pays du nord de l'Europe continentale. Ces pays de plaines se situent au cœur des aires de distribution des espèces eurosibériennes, ce qui limite la perte de ces espèces. Inversement, ces habitats sont trop froids pour les espèces méditerranéennes et trop secs pour les espèces Atlantiques. Cependant, il est très probable qu'en utilisant des scénarios pour 2100, ce front écotone de turnover d'espèces se propage vers le nord de l'Europe et commence à toucher l'Allemagne, mais ceci reste purement hypothétique.

Ces tendances, qui sont les conclusions les plus probables aux objectifs proposés, compte tenu de toutes les précautions méthodologiques prises (variation de modèles et scénarios puis analyse de consensus), rejoignent assez fortement les études montrant les grands couloirs de colonisations des espèces durant les époques glaciaires - interglaciaires qui se situaient dans ces mêmes zones écotonales (Taberlet *et al.* 1998).

Toutefois, une nuance se doit d'être rappelée. Ces taux de turnover n'impliquent que des changements d'espèces et pas de changements d'ordre fonctionnel. Une région fortement perturbée spécifiquement ne subira pas forcément de fort changement fonctionnel, ce qui en retour ne modifiera pas forcément les processus des écosystèmes.

4 Conclusion

Cette dernière partie a montré l'utilité des modèles statiques pour évaluer quel pourrait être l'impact des changements climatiques sur les espèces d'une part, et sur la diversité européenne d'autre part. Malgré l'accumulation d'incertitudes et de biais relatifs à l'utilisation de modèles, au manque de fiabilité des atlas européens et à l'échelle grossière utilisée, des patrons généraux peuvent être dérivés :

1 - **Sensibilité des espèces** : Les propriétés des niches des espèces et leur type chorologique associés à l'exposition aux changements sont des indicateurs fiables de sensibilité. Indépendamment de leur capacité de dispersion et de colonisation, les espèces européennes marginales sur le côté froid du gradient de température sud-nord pourraient être fortement sensibles compte tenu de leur faible amplitude de niche et de leur forte exposition à un réchauffement climatique (ex : espèces boréoalpines). Inversement, les espèces marginales du côté des fortes températures sont susceptibles de gagner des futurs habitats potentiels en direction du nord de l'Europe (ex : espèces méditerranéennes avec des amplitudes de niche

moyennes ou grandes). Les espèces non marginales et peu tolérantes sont elles aussi menacées si elles sont dans une zone exposée à un changement non adéquat pour elles (ex : espèces eurosibériennes).

2 - Sensibilité géographique :

Incertain des projections : La mise en place d'une analyse portant sur 2600 espèces, en utilisant 4 modèles différents (ANN, CTA, GAM, GLM), 2 méthodes de transformation des probabilités en présence-absence (courbe ROC et Kappa), et 5 scénarios de changements climatiques (HadCM3 A1-A2-B1-B2 ; Csrio2 A2) a permis de mettre en évidence que l'incertitude liée à l'utilisation d'un modèle particulier était la principale source de variabilité et troublait les conclusions liées à l'emploi de différents scénarios. Ceci montre que l'utilisation *a priori* d'un type de modèle particulier associée à une méthode donnée de transformation des probabilités de présence en présence-absence, peut biaiser les conclusions relatives à la sensibilité géographique. Cette étude renforce le besoin de choisir le meilleur modèle par espèce pour pouvoir projeter ses futurs habitats potentiels.

Sensibilité des pools régionaux d'espèces par pixel et conséquences écologiques : Après la prise en compte de l'incertitude méthodologique (analyse de consensus) et le choix du couple (modèles x méthode seuil) optimisé, les projections de la sensibilité des pools régionaux d'espèces ont pu être analysées suivant une série de scénarios de changements climatiques (HadCM3 A1-A2-B1-B2).

En moyenne, sous le scénario A1 (économique et global), la diversité européenne devrait être fortement perturbée de manière hétérogène. Le scénario B2 (environnemental et régional) est lui aussi assez perturbateur, tandis que les scénarios A2 (économique et régional) et B1 (environnemental et global) sont les moins perturbants pour la diversité sur une moyenne de changement entre 2020 et 2050. D'un point de vue biogéographique, les conclusions changent assez fortement en fonction des scénarios. Si A1 était en moyenne le plus drastique, il l'est seulement sur la région continentale, dominante en Europe. A l'inverse, se sont sous le scénario B2 que les régions alpine, boréoalpine et pannonienne sont les plus menacées, certainement dû au fait que l'exposition aux changements de ces régions est plus importante sous B2 que sous les autres scénarios.

Les taux d'extinctions discutés sont potentiels car il est impossible de prédire exactement comment vont répondre chacune des espèces modélisées, entre une extinction progressive localement, une plasticité permettant à l'espèce de subsister ou une sélection naturelle peu probable compte tenu de la vitesse actuelle des changements. Toutefois, les données paléoécologiques nous suggèrent que les espèces, malgré un certain temps de réponse, ont eu

tendance à suivre la direction des changements climatiques passés (Prentice *et al.* 1991; Graham 1992; Webb 1992). Implicitement les taux de colonisation dépendent d'une part des capacités de dispersion et de colonisations des espèces et d'autre part de la réaction des écosystèmes et des communautés originales à ces nouveaux arrivants (Lavorel & Garnier 2002; Higgins *et al.* 2003). Dans un site donné, si certaines espèces en régression disparaissent et d'autres parviennent à coloniser, il devrait y avoir une redéfinition des communautés due en partie à des modifications des interactions biotiques (compétitrices et facilitatrices). Si ces changements spécifiques impliquent des changements fonctionnels, il pourrait y avoir des modifications de certains processus écosystémiques. Donc même si les taux prédits ne sont pas exacts, ils traduisent un degré de perturbation spécifique pouvant engendrer des remaniements à l'intérieur des communautés existantes et la création de nouvelles communautés et écosystèmes.

Indépendamment du scénario de changements climatiques, les projections ont mis en évidence une zone où les taux de colonisation et d'extinction locales sont très importants impliquant un fort turnover spécifique. Cette zone correspond sensiblement à l'écotone entre les régions biogéographiques méditerranéennes et eurosibériennes (Figure III.11). Même s'il est très difficile de prédire exactement quelles seront les réponses des espèces dans ces zones, un fort turnover implique des changements de communautés, des modifications dans les écosystèmes originaux et la création de nouveaux écosystèmes, le tout générant une modification des habitats. Cette zone de fort turnover, impliquant des migrations potentielles d'espèces (processus de régression ou de colonisation), est similaire à celle observée durant les périodes de réchauffement post-glaciaire de l'Holocène (Taberlet *et al.* 1998; Blondel & Aranson 1999). Cette zone de brassage spécifique entre les espèces à préférence sèche et chaude et les espèces à d'habitats froids et humides a été observée durant les périodes successives de réchauffement et de refroidissement dans les temps géologiques (Blondel & Aranson 1999). De la même manière que suggéré ici, les montagnes comme les Alpes, les Pyrénées et les Balkans et autres montagnes basses (ex : Cévennes) ont joué un rôle de barrières aux couloirs de migrations, mais aussi un rôle de refuges pour certaines espèces à tendances eurosibériennes durant le réchauffement post-glaciaire. Ces zones refuges ont permis une lente spéciation et différenciation génétique et sont d'ailleurs les sites actuels avec un fort taux d'endémisme en Europe (Jansson 2003).

CONCLUSIONS ET PERSPECTIVES

La biogéographie prédictive a pris son essor cette dernière décennie notamment grâce à la puissance des nouveaux outils informatiques, les nouvelles évolutions des systèmes d'informations géographiques et la compréhension des mécanismes de distributions des espèces à l'échelle continentale. Cependant, la biogéographie prédictive couplée à la macroécologie sont des disciplines relativement jeunes, qui doivent encore s'articuler autour de principes scientifiques généraux. La réticence des écologistes à la macroécologie provient du fait que cette discipline reste très descriptive et que peu de principes ou mécanismes généraux ont été extrapolés. Ce champ disciplinaire ne doit cependant pas être occulté par l'écologie générale car il permet d'étudier la réponse des organismes à des macro-variables (Brown 1995), d'analyser les relations taille des aire de distributions - taille des organismes ('range size - body size', Gaston 2003), ou encore d'analyser et évaluer les réponses des organismes aux changements de milieu sur de vastes étendues (Gaston 2000). Lawton (1999) décrit la macroécologie comme « la recherche de structures statistiques dans les types, les distributions, les abondances et la richesse des espèces, de l'échelle locale à globale, ce qui permet les développements et les tests d'explications théoriques sous-jacentes à ces patrons ». En ce sens, la macroécologie doit aider à comprendre les patrons de distributions des espèces et permettre d'appréhender l'impact des changements de milieu sur ces distributions. Si l'étude des communautés et des populations permet d'étudier la réponse d'un groupe d'espèces et de toutes les interactions possibles avec leur milieu en termes évolutifs ou démographiques, la macroécologie permet de déterminer une image plus générale où un ordre statistique émerge (Blackburn & Gaston 1998; Gaston & Blackburn 1999; Maurer 2000). Bien que l'écologie des communautés puisse aider à comprendre la réponse de quelques espèces aux changements globaux (Lavorel & Garnier 2002), la complexité des relations émergentes ne permet pas de faire de prédictions généralisables et réalistes (Lawton 1999).

Dans cette thèse, la voie macroécologique a été utilisée pour étudier l'impact potentiel des changements globaux sur la distribution des espèces et indirectement sur la diversité européenne. Cette simplification a permis de s'affranchir des relations complexes qui régissent les communautés (compétition, interactions entre niveaux trophiques) et des paramètres difficiles à estimer sur un grand nombre d'espèces tels que la dispersion ou la colonisation, et ainsi de pouvoir déterminer des grands patrons de réponses à travers les différents taxons étudiés. Cette approche ne permet cependant pas de prédire avec exactitude ce qui va se passer dans un futur proche, mais permet de dresser quelques patrons généraux décrivant l'impact probable des

changements globaux sur les espèces. Mon travail, qui se situe à l'interface entre la biogéographie, l'écologie et la modélisation biostatistique a permis de mettre en évidence :

- qu'un effort reste à faire sur l'harmonisation et la fiabilité des techniques de modélisation employées. L'incertitude liée à l'utilisation de méthodes diverses est trop importante pour dresser des conclusions fiables dans le futur,
- que l'exposition géographique des espèces reste le paramètre fondamental de la sensibilité des espèces,
- que les différents chorotypes européens ne sont pas égaux devant les changements globaux, dépendant de leurs propriétés écologiques et de leur exposition.
- et que selon les projections futures, les régions biogéographiques aussi bien que les pays n'ont pas la même sensibilité aux différents scénarios de changements globaux.
- une zone écotone entre les régions biogéographiques méditerranéenne, atlantique et eurosibérienne où le turnover d'espèce devrait être important, synonyme de fortes perturbations à l'intérieur des habitats, communautés et potentiellement sur le fonctionnement des écosystèmes.

Je ne reprendrai pas ici l'ensemble des résultats obtenus au cours des différentes études et analyses développées durant cette thèse, mais ferai plutôt un état des lieux des résultats qui sont nouveaux par rapport au traitement des objectifs et questions liées aux changements globaux, et de ce qui reste à faire pour améliorer la compréhension des phénomènes et la fiabilité des projections.

1. Considérations écologiques et biogéographiques

Comment les changements globaux récents et futurs influent-ils sur la distribution des espèces et indirectement sur la diversité spécifique ? Malgré la grande incertitude liée au choix de la méthode de modélisation empirique, mes résultats montrent que d'une part les espèces répondent de manière variée aux changements globaux, et que d'autre part cette réponse est fonction des propriétés écologiques des espèces et de leur exposition géographique. Toutes les espèces n'ont donc pas les mêmes chances vis à vis des changements. A n'importe quelle échelle, les propriétés écologiques des espèces en termes de position et d'amplitude de niche réalisée dans la zone d'étude sont facilement mesurables et permettent d'évaluer la sensibilité relative des espèces aux changements de milieu. Nous avons vu qu'à échelle européenne, les

propriétés de niches des espèces peuvent être décrites de manière bioclimatique et permettent de déterminer par exemple que les espèces boréoalpines sont fortement sensibles aux changements compte tenu de leur marginalité et de leur faible amplitude bioclimatique ainsi que de leur forte exposition à un réchauffement. Les propriétés de niche des espèces ont ainsi pu être corrélées avec les projections des futurs habitats potentiels des espèces, permettant de faire le lien entre des propriétés écologiques (spécialistes vs généralistes, marginales vs ubiquistes) et des projections mathématiques. L'utilisation des propriétés des niches dans la zone d'étude pour déterminer la sensibilité des espèces à un changement pourrait aussi être appliquée à d'autres échelles et d'autres variables. Il est généralement accepté qu'à l'échelle du paysage le climat ne joue qu'un rôle secondaire par rapport aux activités humaines, la couverture végétale ou la géologie. Les propriétés de niche des espèces peuvent donc être déduites de ces variables pour déterminer leur position et surtout leur amplitude d'habitats utilisés par rapport aux changements attendus dans la zone d'étude. Cette stratégie permettrait, comme nous l'avons fait dans cette thèse, de confronter la sensibilité des espèces, déterminée sur la base unique de leurs propriétés écologiques, avec les prédictions des modèles.

D'une manière plus générale, cette thèse a aussi permis de délimiter la région en Europe qui pourrait être soumise à de fortes extinctions et colonisations d'espèces, et donc à une forte modification des communautés et potentiellement du fonctionnement des écosystèmes si des changements fonctionnels se produisent. Cette région qui représente en gros les écotones des grandes régions biogéographiques européennes a déjà été le théâtre de profonds changements durant les périodes glaciaires et inter- glaciaires (Taberlet *et al.* 1998; Blondel & Aranson 1999). En effet cette zone correspond à l'interface entre les bornes sud des aires de répartition des espèces eurosibériennes et les bornes nord des espèces méditerranéennes. Durant les périodes inter-glaciaires de réchauffement, les espèces eurosibériennes et boréales avaient déjà été en recul dans le sud de l'Europe et emprunté les couloirs de migration correspondant à cette zone de fort turnover (Taberlet *et al.* 1998). Durant cette période, les zones barrières comme les Pyrénées et les Alpes avaient créé des zones refuges pour un grand nombre d'espèces et ainsi maintenu une forte diversité dans ces régions de brassage (Taberlet & Cheddadi 2002). Ces mêmes zones barrières se retrouvent dans la zone de fort turnover décrite ici, et on peut supposer qu'elles devraient encore jouer un rôle de refuges pour certaines espèces eurosibériennes et peut-être permettre la création d'une nouvelle forme endémisme par spéciation (Jansson 2003). Toutefois, notre analyse ne prend en compte qu'un faible pourcentage de la diversité européenne réelle. En effet, la flore de la région méditerranéenne est fortement sous échantillonnée dans l'Atlas Florae Europaeae où seulement 15% de la flore totale est représentée. Ce manque est crucial car il

pourrait évidemment changer les conclusions sur cette région. A l'inverse, les régions nordiques de l'Europe sont assez bien représentées et ceci dans les quatre atlas, faisant que les projections de diversité peuvent être jugées assez fiables.

Il est bien sûr impossible de dire si les perturbations projetées seront effectivement réalisées car cela dépend de paramètres stochastiques et biologiques non pris en compte. Premièrement, les taux d'extinction locale prédits sont dépendant d'une balance entre les taux d'extinction réels et la plasticité phénotypique des espèces, tandis que les taux de colonisation sont fortement dépendants de la capacité de dispersion et de colonisation des espèces (Higgins *et al.* 2003). Deuxièmement, les changements de distribution d'espèces vont modifier les interactions interspécifiques et donc les niches réalisées des espèces. Par exemple, l'extinction d'une espèce dominante permettra par exemple à une espèce jadis dominée de s'étendre, et donc de modifier sa niche réalisée, ce qui modifiera en retour les communautés et certainement le fonctionnement de certains écosystèmes (Chapin III *et al.* 2000; Lavorel & Garnier 2002). De même l'arrivée de nouveaux compétiteurs pourraient modifier les hiérarchies compétitives et facilitatrice et donc influencer fortement sur la productivité des écosystèmes (Loreau 2000; Nijss & Roy 2000; Mouquet *et al.* 2002). Malgré tout, on peut supposer que c'est dans ces régions que le risque de turnover ou d'extinction est le plus important et que des efforts de conservation doivent être menés.

2. Considérations méthodologiques

Que ce soit en macroécologie ou plus particulièrement en biogéographie prédictive, l'importance des modèles et l'interprétation qu'en fera le modélisateur est fondamentale. Récemment, divers articles ont publié des résultats alarmants et pessimistes sur l'impact des changements climatiques sur la biodiversité (Bakkenes *et al.* 2002; Erasmus *et al.* 2002; Peterson *et al.* 2002). Ces analyses réalisées sur un grand nombre d'espèces végétales ou animales ont montré la vulnérabilité géographique des zones étudiées en fonction d'un scénario de changements climatiques et en utilisant un seul modèle statique. Par exemple, Peterson *et al.* (2002) utilisent des algorithmes génétiques, Bakkenes *et al.* (2002) utilisent des modèles linéaires généralisés et Erasmus *et al.* (2002) un type d'enveloppe bioclimatique. Un problème se pose sur la comparabilité des résultats et surtout sur la fiabilité de ces projections. D'une part utiliser un seul scénario de changement climatique pour déterminer la sensibilité d'un grand nombre d'espèces sur de vastes étendues c'est s'affranchir de l'incertitude climatique. En effet, différents GCM existent avec différents scénarios d'émission ce qui permet de donner une

amplitude dans les projections. Pourquoi utiliser une seule et unique combinaison des changements futurs sachant l'incertitude qui existe dans la construction des modèles climatiques (Watson & et. 2001) ? D'autre part, nous avons vu qu'il existe une grande variabilité dans les projections d'habitats potentiels entre modèles statiques utilisés. En effet, de nombreuses études ont comparé la puissance prédictive des modèles statiques sur les conditions actuelles (Franklin 1998; Guisan *et al.* 1999; Manel *et al.* 1999; Moisen & Frescino 2002; Thuiller *et al.* 2003), mais peu ont analysé si les différences observées sur les prédictions actuelles étaient conservées, exacerbées ou diminuées sur les projections dans le futur. Dans cette thèse j'ai montré que les différences étaient dans de trop nombreux cas exacerbées, rendant incertaines les projections dans le futur. Si deux modèles ont la même puissance prédictive sur les données actuelles mais qu'ils prédisent deux distributions très différentes dans le futur, lequel croire ? Ceci pose bien évidemment la question de la validité et la pertinence des études présentées au début du paragraphe.

Ici, nous proposons une approche complètement différente qui serait de prendre en compte l'incertitude méthodologique. Cette approche utilise une gamme de modèles statiques (non paramétriques *vs* paramétriques, statistiques *vs* basés sur les règles) et permet de choisir le modèle qui dans le futur résume le plus d'information commune à tous les autres modèles. En effet, les modèles projettent souvent des distributions similaires, mais différent sur les limites d'aires de distribution. Choisir un modèle qui maximise les zones où les modèles sont similaires et minimise les zones où ils sont discordants permet d'avoir une approche conservatrice. Ceci ne veut cependant pas dire que ce modèle sera le meilleur dans le futur, mais qu'actuellement il est le meilleur compromis. Cela suppose donc que le modélisateur reconnaisse l'incertitude liée à ces projections, car même si avec cette méthode elle est diminuée, elle demeure toujours.

3. Vers une approche basée sur les groupes de réponse

Nous avons vu que les propriétés de niche des espèces permettaient d'évaluer leur sensibilité aux changements climatiques. Des espèces qui ont des propriétés de niche similaires auront des réponses similaires aux changements de milieu. D'un autre côté, le concept de Type Fonctionnel Végétal (PFT) suggère que les espèces puissent être groupées selon leurs réponses communes à l'environnement et/ou selon leurs effets sur les processus des écosystèmes (Gitay & Noble 1997; Lavorel & Garnier 2001, 2002). Des analyses récentes ont mis en évidence que des espèces avec des réponses similaires au milieu partageaient des traits biologiques similaires: Parmi ces traits, certains dits 'softs', facile à mesurer sur un grand nombre d'espèces mais qui ne

Encadré C.1

Objectif : mettre en relation les propriétés écologiques des espèces suivant des gradients bioclimatiques et les traits biologiques qui les composent.

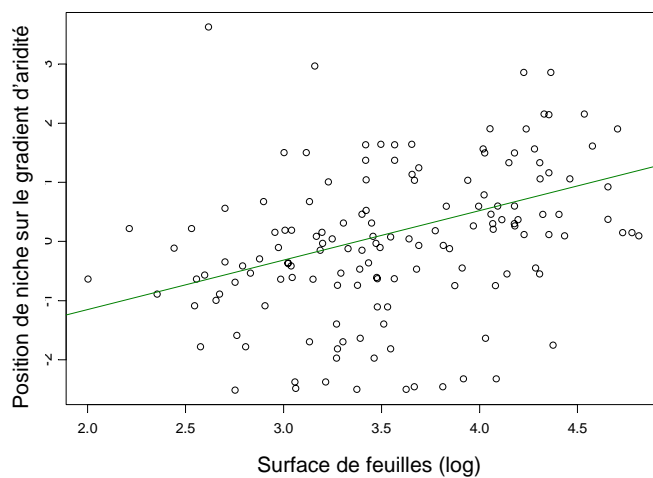
Données espèces : 88 espèces du genre *Leucadendron* endémique de la région floristique du Cap (Williams 1972). Ce sont des arbustes sempervirents et dioïques. La distribution spatiale des espèces a été enregistrée durant un échantillonnage massif de la famille des Protea en Afrique du Sud (Rebello 1992). La présence des espèces a été notée sur plus de 30000 sites à travers la région et est disponible sous format SIG (Rebello 2001). Les traits morphologiques, phénologiques et de dispersion ont été extraits de la base de données rattachée à cet échantillonnage (Rebello 2001).

Données bioclimatiques : les mêmes variables que dans l'encadré II.2.

Propriété des niches : Nous avons utilisé l'analyse OMI (Doledec *et al.* 2000) déjà présentée précédemment. OMI permet de distinguer trois grands gradients climatiques expliquant 94% de la variabilité totale : gradient d'aridité - saisonnalité des précipitations - température.

Les traits morphologiques, phénologiques et de dispersion ont ensuite été mis en relation avec les propriétés des niches sur chacun des gradients.

Figure C.1 : Relation position de niche sur le gradient d'aridité avec la surface de feuille des espèces de *Leucadendron* ($R^2=22,3\%$).



sont pas explicitement liés à des mécanismes fonctionnels spécifiques sont principalement liés à la réponse des espèces aux modifications du climat, aux régimes de perturbations et à l'augmentation de CO₂ (ex : forme, surface et taille de feuilles, surface spécifique des feuilles, taux de croissance relatif, hauteurs de plantes (Bond & Midgley 1988; Cornelissen 1999; Poorter & Garnier 1999; Niinemets 2001; Cornelissen *et al.* 2003).

Il y a donc un point de convergence sur la réponse similaire des espèces au milieu avec d'une part leurs propriétés de niche et d'autre part les traits biologiques de réponse. Si l'on peut réunir ces deux jeux d'informations spécifiques (niches + traits), on doit pouvoir définir des groupes de réponses assez stables aux changements globaux considérés. En effets, les modèles de végétation mécanistes à grande échelle (Woodward & Cramer 1996; Smith *et al.* 1997) ont groupé des plantes *a priori*, sur la base de la connaissance de leur fonction ou des corrélations observées parmi des traits fonctionnels qui devraient permettre de prédire les changements sur les écosystèmes à partir des changements projetés dans la composition des plantes en réponse aux changements globaux (Woodward & Cramer 1996; Smith *et al.* 1997). Ces modèles utilisent principalement des types fonctionnels tels que les herbacées, conifères ou des plantes tropicales en C4 (Sitch *et al.* 2003). Toutefois, compte tenu de la variabilité de chorotypes à l'intérieur des Conifères, on peut douter de la fiabilité d'une telle classification.

Inversement, si des espèces avec des propriétés de niche similaires partagent des traits de réponse similaires, alors nous pourrions former des groupes de réponse aux changements globaux et avoir des indices de sensibilité non contraints par l'exposition géographique des espèces. En effet, si des espèces marginales d'un côté d'un gradient X partagent des traits de réponses similaires, alors le couple Marginalité x Traits permet de ne plus prendre en compte la position géographique de l'espèce car les deux informations permettraient de le faire. Il serait donc ainsi aisé de cartographier ces groupes ainsi formés et de prédire leur distribution à l'aide de modèles statiques comme ceux utilisés dans ce manuscrit.

Dans une analyse préliminaire réalisée en Afrique du Sud sur 88 espèces du genre *Leucadendron*, nous avons montré que la position et l'amplitude de niche des espèces étaient corrélées avec des traits morphologiques (surface de feuille), phénologiques (début et fin de floraison) et de dispersion (agent de dispersion) (Thuiller *et al.* in press, Annexe 4.2) (Encadré C.1). Par exemple, la position de la niche des espèces sur le gradient d'aridité est corrélée avec la surface des feuilles, démontrant que la marginalité des espèces sur ce côté du gradient est exprimée en terme de réduction de surface foliaire (Figure C.1). Il y a donc bien une relation forte entre les propriétés de niches des espèces et les traits de réponses sélectionnés. Des études plus approfondies seront nécessaires pour analyser en Europe ce couple niches - traits de

réponses permettant la construction de groupes de réponses aux changements globaux et ainsi d'étudier leur sensibilité aux changements globaux.

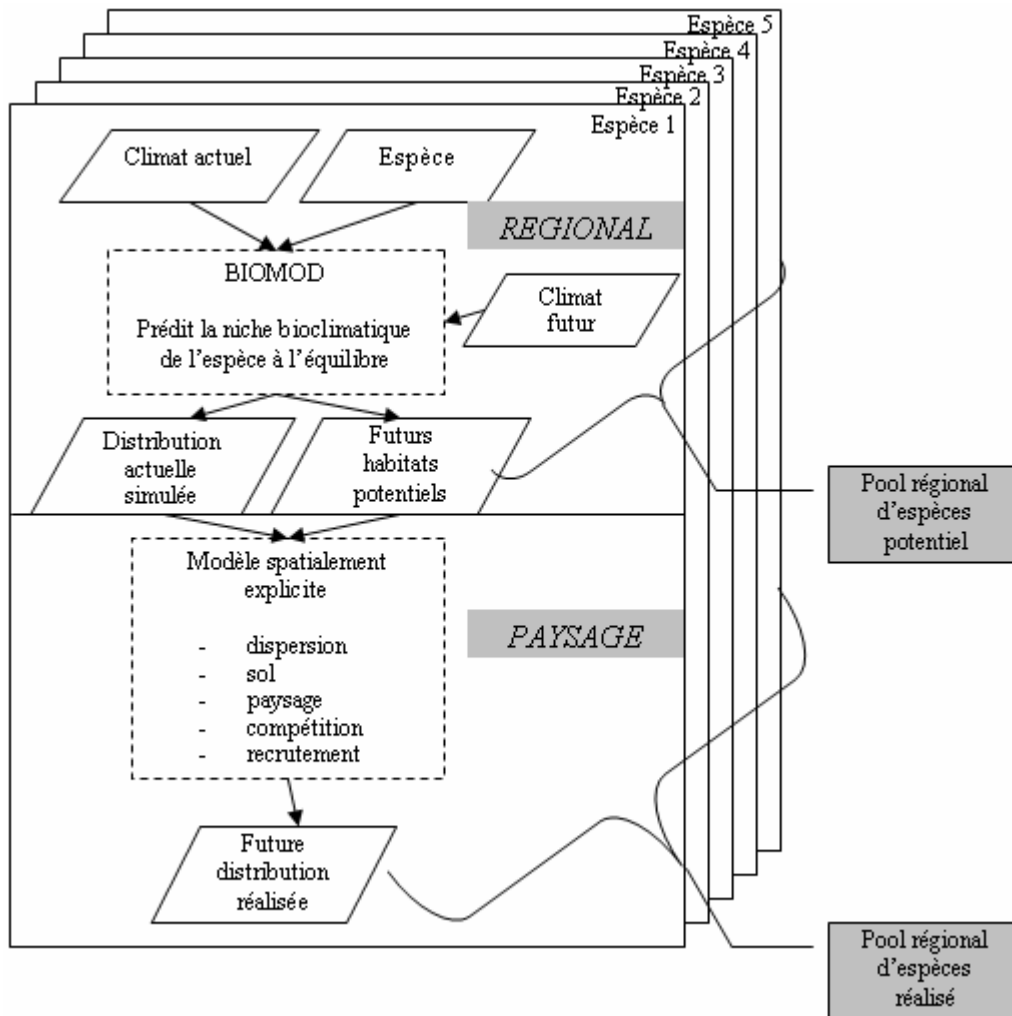
4. Vers une approche multi-échelle spatialement explicite

Dans cette thèse j'ai démontré l'utilité des modèles statiques pour évaluer quel pourrait être l'impact des changements climatiques sur les espèces. Cependant, par leur approche statique, ces modèles ne donnent qu'une image fixe à un temps t , sans pouvoir prendre en compte les paramètres de dispersion ou de colonisation des espèces, de la structure du paysage et des changements d'utilisation des terres. Nous avons vu que l'utilisation des terres à échelle européenne était partiellement contrôlée par le climat et qu'inclure cette variable dans des modèles bioclimatiques n'était pas nécessaire. Pourtant il est évident qu'à échelle locale la modification de l'utilisation des terres perturbe les habitats et provoquent des extinctions, en particulier chez les espèces spécialistes ou rares. Dans une analyse réalisée pour l'Hérault, nous avons montré que l'urbanisation croissante et l'intensification de l'agriculture sur les 100 dernières années avaient provoqué la disparition d'espèces rares dans la région d'étude (Lavergne *et al.* in prep, Annexe 4.1). Il y a de fortes preuves empiriques que divers facteurs influencent la distribution des espèces d'une manière hiérarchique et spatiale (Willis & Whittaker 2002).

Pour évaluer d'une manière plus fiable l'impact des changements globaux sur un certain nombre d'espèces, il semble nécessaire de prendre en compte tous ces déterminants agissant à diverses échelles et principalement la migration des espèces. Une approche qui semblerait pertinente serait de combiner des modèles dynamiques de paysages avec des modèles statiques bioclimatiques. Ceci procurerait des capacités de prédictions accrues, incluant la simulation de la compétition, des régimes de perturbation, de la dispersion et des effets du changement d'utilisation des terres. Ceci permettrait de prédire les futures distributions potentielles des espèces et non pas les futurs habitats potentiels comme présentés dans cette thèse.

En effet, si la distribution actuelle d'une espèce et les futurs habitats potentiels prédits en 2050 ou 2100 par BIOMOD sont disjoints, il est évident qu'il manque les mécanismes dynamiques nécessaires pour reproduire les conditions à non-équilibre où les limites actuelles et potentielles de l'espèce ne se recoupent pas, ainsi que les processus tels que la persistance et l'extinction dans des sites non adéquats, la dispersion vers de nouveaux sites et les phases d'établissement des espèces dans de nouvelles conditions. Iverson *et al.* (1999) utilisent des arbres de régression pour modéliser la niche réalisée de 60 espèces d'arbres, et des automates

Figure C.2 : Schéma représentatif de l'approche multi-échelle pour modéliser la dynamique d'espèces (ou groupes de réponse) dans un contexte de changement global.



cellulaires (AC) pour simuler les processus de migration à travers le paysage. Les auteurs ont volontairement minimisé les hypothèses et approximations sur des variables biologiques importantes (ex : établissement des graines, mortalité juvénile, compétition interspécifique) en favorisant l'hypothèse simple que les espèces peuvent maintenir une migration moyenne de cinquante kilomètres par an dans des régions entièrement forestière. Ce taux qui est la borne inférieure de ceux observés durant l'Holocène (Webb 1992) sont ralentis en fonction de la disponibilité des habitats (ex : fragmentation). Bien que très simpliste, ce type de modèles avec des règles simples permet de prendre en compte les processus de dispersion (courte et longue distance) à travers le paysage.

Des modèles un peu plus complexes pourraient être utilisés comme ceux développés sous la plateforme LAMOS (Lavorel *et al.* 2000) qui permettent de prendre en compte différents niveaux de complexités des processus impliqués dans la dynamique du paysages (succession, perturbation, dispersion). L'idéal serait donc de construire la niche bioclimatique potentielle d'une espèce (ou groupe de réponse) à échelle régionale au temps t , à l'aide de modèles statiques comme BIOMOD (Figure C.2). Ensuite des scénarios de changements climatiques seraient appliqués au temps $t+1$ (fonction des espèces et de la disponibilité des données) pour modéliser la répartition des futurs habitats potentiels (comme réalisé dans cette thèse). Dans chaque pixel, la dynamique de l'espèce sera ensuite modélisée, fonction des propriétés intrinsèques de l'espèce (dispersion, régénération, recrutement, compétition) et de la structure du paysage (fragmentation, perturbation, succession, changement d'utilisation des terres). Cette approche à échelle du paysage permettra de modéliser la future distribution réalisée de l'espèce modélisée au temps $t+1$. Le cycle peut ensuite continuer au temps $t+2$ (Figure C.2). A ce niveau basé sur l'espèce ou groupe de réponse peut se rajouter une vue globale du pool régional d'espèces ou de groupes de réponse par pixel. Cette méthode permettrait de modéliser le pool régional potentiel dans un temps futur et le pool effectivement réalisé. Il serait ensuite aisé de dériver des conclusions sur l'impact des changements globaux sur les communautés (turnover) et sur les écosystèmes étudiés (productivité).

La qualité prédictive du couplage pourra être mesurée par une approche rétrospective en modélisant les changements de distribution de quelques espèces arbres (genre *Quercus* et *Fagus*) durant la déglaciation post-Pléistocène. Ces projections seront ensuite comparées aux cartes de distribution issues des données polliniques.

Pour conclure, la méthode qui me semble le plus prometteuse est donc un couplage des modèles dynamiques de paysage et des modèles statiques de distributions à grandes échelles, permettant ainsi de travailler aux interfaces de différentes sous-disciplines de l'écologie.

BIBLIOGRAPHIE

- Ahn, C.H., Tateishi R. (1994) Development of a global 30-minute grid potential evapotranspiration data set. *Journal of Japanese Soc. Photogram. Rem. Sens.* **33**, 12-21.
- Allen, A.P., O'Connor R.J. (2000) Interactive effects of land use and other factors on regional bird distributions. *Journal of Biogeography* **27**, 889-900.
- Anderson, P.M., Brubaker L.B. (1994) Vegetation history of northcentral Alaska - A mapped summary of late Quaternary pollen data. *Quaternary Science Reviews* **13**, 71-92.
- Anon (1999) *S-PLUS 2000 User's Guide*. MathSoft, Seattle, WA.
- Araújo, M.B., Thuiller W., Williams P.H., Reginster I. (soumis) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology & Biogeography*.
- Araújo, M.B., Williams P. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation* **96**, 331-345.
- Araújo, M.B., Williams P.H., Turner A. (2002a) A sequential approach to minimise threats within selected conservation areas. *Biodiversity and Conservation* **11**, 1011-1024.
- Araújo, M.B., Williams S.E., Fuller R.J. (2002b) Dynamics of extinction and the selection of nature reserves. *Proceedings of the Royal Society London B*.
- Austin, M.P. (1998) An ecological perspective on biodiversity investigations: examples from Australian Eucalypt forests. *Ann. Missouri. Bot. Gard.* **85**, 2-17.
- Austin, M.P. (2002) Spatial Prediction of Species Distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* **157**, 101-118.
- Austin, M.P., Gaywood M.J. (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., Meyers J.A. (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., Nicholls A.O., Doherty M.D., Meyers J.A. (1994) Determining species response functions to an environmental gradient by means of a beta-function. *Journal of Vegetation Science* **5**, 215-228.
- Austin, M.P., Nicholls A.O., Margules C.R. (1990) Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs* **60**, 161-177.
- Austin, M.P., Smith T.M. (1989) A new model for the continuum concept. *Vegetatio* **83**, 35-47.
- Bachelet, D., Neilson R.P., Hickler T., Drapek R.J., Lenihan J.M., Sykes M.T., Smith B., Sitch S., Thonicke K. (2003) Simulating past and future dynamics of natural ecosystems in the United States. *Global Biogeochemical Cycles* **17**, 14.11-14.21.
- Bairlein, F., Winkel W. (2001) Birds and climate change. In: *Climate of the 21st Century: Changes and Risks* (eds Lozan JL, Graßl H, Hupfer P), pp. Wissenschaftliche Auswertungen, Hamburg.
- Bakkenes, M., Alkemade R.M., Ihle F., Leemans R., Latour J.B. (2002) Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology* **8**, 390-407.

- Baquero, R.A., Telleria J.L. (2001) Species richness, rarity and endemism of European mammals: a biogeographical approach. *Biodiversity and Conservation* **10**, 29-44.
- Barbosa, A.M., Real R., Olivero J., Vargas J.M. (2003) Otter (*Lutra lutra*) distribution modelling at two resolution scales suited to conservation planning in the Iberian Peninsula. *Biological Conservation* **114**, 377-387.
- Bartlein, P.J., Prentice I.C., Webb T. (1986) Climatic response surfaces from pollen data for some eastern North American taxa. *Journal of Biogeography* **13**, 35-57.
- Bascompte, J., Solé R.V. (1996) Habitat fragmentation and extinction thresholds in spatially explicit models. *Journal of Animal Ecology* **65**, 465-473.
- Beerling, D.J., Huntley B., Bailey J.P. (1995) Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surface. *Journal of Vegetation Science* **6**, 269-282.
- Beerling, D.J., Kelly C.K. (1997) Stomatal density responses of temperate woodlands plants over the past seven decades of CO₂ increase: a comparison of Salisbury (1927) with contemporary data. *American Journal of Botany* **84**, 1572-1583.
- Benton, T.G., Vickery J.A., Wilson J.D. (2003) Farmland biodiversity: is habitat heterogeneity the key? *TREE* **18**, 182-188.
- Bio, A.M.F., Alkemade R., Barendregt A. (1998) Determining alternative models for vegetation response analysis: a non-parametric approach. *Journal of Vegetation Science* **9**, 5-16.
- Bio, A.M.F., de Becker P., de Bie E., Huybrechts W., Wassen M. (2002) Prediction of plant species distribution in lowland river valleys in Belgium: modelling species response to site conditions. *Biodiversity and Conservation* **11**, 2189-2216.
- Blackburn, T.M., Gaston K.J. (1998) Some methodological issues in Macroecology. *The American Naturalist* **151**, 68-83.
- Blondel, J., Aranson J. (1999) *Biology and wildlife of the Mediterranean region*. Oxford University Press, New-York.
- Bolliger, J., Kienast F., Bugmann H. (2000) Comparing models for distributions: concept, structures, and behavior. *Ecological Modelling* **134**, 89-102.
- Bolòs, O.d., Vigo J., Masalles R.M., Ninot J.M. (1990) *Flora manual dels Països Catalans*. Pòrtic, Barcelona.
- Bond, W.J., Midgley J.J. (1988) Allometry and sexual differences in size. *The American Naturalist* **131**, 909-910.
- Bradley, N.L., Leopold A.C., Ross J., Huffaker W. (1999) Phenological changes reflect climate change in Wisconsin. *Proceedings of the National Academy of Science, USA* **96**, 9701-9704.
- Breiman, L., Friedman J.H., Olshen R.A., Stone C.J. (1984) *Classification and regression trees*. Chapman and Hall, New York.
- Breitenmoser, U. (1998) Large predators in the Alps: the fall and rise of man's competitors. *Biological Conservation* **83**, 279-289.
- Bridge, S.R.J., Johnson E.A. (2000) Geomorphic principles of terrain organization and vegetation gradients. *Journal of Vegetation Science* **11**, 57-70.
- Brown, D.G. (1994) Predicting vegetation types at treeline using topography and biophysical disturbance variables. *Journal of Vegetation Science* **5**, 641-656.
- Brown, J.H. (1995) *Macroecology*. The University of Chicago Press, Chicago.

- Brown, J.H., Ernest S.K.M., Parody J.M., Haskell J.P. (2001) Regulation of diversity: maintenance of species richness in changing environments. *Oecologia* **126**, 321-332.
- Brubaker, L.B. (1986) Responses of tree populations to climatic change. *Vegetatio* **67**, 119-130.
- Brzeziecki, B., Kienast F., Wildi O. (1995) Modelling potential impacts of climate change on the spatial distribution of zonal forest communities in Switzerland. *Journal of Vegetation Science* **6**, 257-268.
- Budyko, M.I. (1974) *Climate and life*. Academic Press, New-York.
- Busby, J.R. (1991) BIOCLIM - a bioclimate analysis and prediction system. In: *Nature Conservation: Cost Effective Biological Surveys and Data Analysis* (eds Margules CR, Austin MP), pp. 64-68. CSIRO, Canberra, Australia.
- Cain, M.L., Damman H., Muir A. (1998) Seed dispersal and the Holocene migration of woodlands in southern Sweden. *Ecological Monographs* **68**, 325-347.
- Carpenter, G., Gillison A.N., Winter J. (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation* **2**, 667-680.
- Chamberlain, D.E., Fuller R.J. (1999) Density-dependent habitat distribution in birds: issues of scale, habitat definition and habitat availability. *Journal of Avian Biology* **30**, 427-436.
- Chamberlain, D.E., Wilson A.M., Browne S.J., Vickery J.A. (1999) Effects of habitat type and management on the abundance of skylarks in the breeding season. *Journal of Applied Ecology* **36**, 856-870.
- Chapin III, S.F., Zavaleta E.S., Eviner V.T., Naylor R.L., Vitousek P.M., Reynolds H.L., Hooper D.U., Lavorel S., Sala O.E., Hobbie S.E., Mack M.C., Diaz S. (2000) Consequences of changing biodiversity. *Nature* **405**, 234-242.
- Chuine, I., Cambon G., Comtois P. (2000) Scaling phenology from the local to the regional levels: advances from species-specific phenological models. *Global Change Biology* **6**, 1-11.
- Clark, J.S. (1998) Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *The American Naturalist* **152**, 204-224.
- Clark, L.A., Pregibon D. (1992) Tree-based models. In: *Statistical Models in S* (eds Chambers JM, Hastie TJ), pp. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California.
- Clarkson, B.R., McGlone M.S., Lowe D.J., Clarkson B.D. (1995) Macrofossils and pollen representing forests of the pre-taupo volcanic eruption (c. 1850 yr bp) era at pureora and benneydale, central north island, new zealand. *Journal of the Royal Society of New Zealand* **25**, 263-281.
- Collingham, Y.C., Wadsworth R.A., Huntley B., Hulme P.E. (2000) Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology* **37**, 13-27.
- Cornelissen, J.H.C. (1999) A triangular relationship between leaf size and seed size among woody species: allometry, ontogeny, ecology and taxonomy. *Oecologia* **118**, 248-255.
- Cornelissen, J.H.C., Lavorel S., Garnier E., Díaz S., Buchmann N., Gurvich D.E., Reich P.B., ter Steege H., Morgan H.D., van der Heijden M.G.A., Pausas J.G., Poorter H. (2003) Handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian Journal of Botany* **51**, 335-308.
- Cowling, R.M., Gibbs Russel G.E., Hoffman M.T., Hilton-Taylor C. (1989) Patterns of plant species diversity in southern Africa. In: *Biotic Diversity in Southern Africa. Concepts and Conservation* (eds Huntley BJ), pp. 19-50. Oxford University Press, Cape Town.

- Cowling, R.M., Lombard A.T. (2002) Heterogeneity, speciation/extinction history and climate: explaining regional plant diversity patterns in the Cape Floristic Region. *Diversity and Distributions* **8**, 163-179.
- Crick, H.Q.P., Sparks T.H. (1999) Climate change related to egg-laying trends. *Nature* **399**, 423-424.
- Croonquist, M.J., Brooks R.P. (1993) Effects of habitat disturbance on bird communities in riparian corridors. *Journal of Soil and Water Conservation* **48**, 65-70.
- CSIR (1999) *South African National Land Cover Database Project. Data Users Manual (Final Report ENV/P/C 98136)*. CSIR, Pretoria, South Africa.
- Davis, A.J., Lawton J.H., Shorrocks B., Jenkinson L.S. (1998) Individualistic species responses invalidate simple physiological models of community dynamics under global environmental change. *Journal of Animal Ecology* **67**, 600-612.
- de Candolle, A.I. (1855) *Géographie botanique raisonnée*. Masson, Paris.
- Doledec, S., Chessel D., Gimaret-Carpentier C. (2000) Niche separation in community analysis: a new method. *Ecology* **81**, 2914-2927.
- Dray, S., Chessel D., Thioulouse J. (In press) Co-inertia analysis and the linking of ecological tables. *Ecology*.
- Duckworth, J.C., Bunce R.G.H., Malloch A.J.C. (2000a) Modelling the potential effects of climate change on calcareous grasslands in Atlantic Europe. *Journal of Biogeography* **27**, 347-358.
- Duckworth, J.C., Bunce R.G.H., Malloch A.J.C. (2000b) Vegetation-environment relationships in Atlantic European calcareous grasslands. *Journal of Vegetation Science* **11**, 15-22.
- Easterling, D.R., Meehl G.A., Parmesan C., Changnon T.R., Karl T.R., Mearns L.O. (2002) Climate extremes: Observations, modeling, and impacts. *Science* **289**, 2068-2074.
- Elith, J., Burgman M.A. (in press) Habitat models for PVA. In: *Population Viability in Plants* (eds Brigham CA, Schwartz MW), pp. Springer-Verlag, New York.
- Elton, C.S. (1927) *Animal ecology*. Sedgwick and Jackson, London.
- Erasmus, B.F.N., Van Jaarsveld A.S., Chown S.L., Kshatriya M., Wessels K.J. (2002) Vulnerability of South African animal taxa to climate change. *Global Change Biology* **8**, 679-693.
- Fitter, A.H., Fitter R.S.R. (2002) Rapid changes in flowering time in British plants. *Science* **296**, 1689-1691.
- Forman, R.T.T., Godron M. (1986) *Landscape ecology*, first edition. John Wiley & Sons, U.S.A.
- Francis, A.P., Currie D.J. (2003) A globally consistent richness-climate relationship for Angiosperms. *The American Naturalist* **161**, 523-536.
- 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.
- Gaston, K.J. (1994) *Rarity*. Chapman and Hall, London.
- Gaston, K.J. (2000) Global patterns of biodiversity. *Nature* **405**, 220-227.
- Gaston, K.J. (2003) *The structure and dynamics of geographic ranges*. Oxford University Press, Oxford.
- Gaston, K.J., Blackburn T.M. (1999) A critic for macroecology. *Oikos* **84**, 353-368.
- Gitay, H., Noble I.R. (1997) What are functional types and how should we seek them? In: *Plant Functional Types: Their Relevance to Ecosystem Properties and Global Change* (eds Smith TM, Shugart HH, Woodward FI), pp. 3-19. Cambridge University Press, Cambridge, UK.
- Gleason, H.A. (1926) The individualistic concept of the plant association. *Bull. Torrey Bot. Club* **53**, 1-20.

- Gottfried, M., Pauli H., Reiter K., Grabherr G. (1995) A fine-scaled model for changes in species distribution patterns of high mountain plants induced by climate warming. *Diversity and Distributions* **5**, 241-251.
- Gracia, C., Burriel J.A., Ibàñez J.J., Mata T., Vayreda J. (2000a) *Inventari Ecològic i Forestal de Catalunya*, Bellaterra.
- Gracia, C., Burriel J.A., Ibàñez J.J., Mata T., Vayreda J. (2000b) *Inventari Ecològic i Forestal de Catalunya*, Bellaterra.
- Gracia, C., Burriel J.A., Ibàñez J.J., Mata T., Vayreda J. (2000c) *Inventari Ecològic i Forestal de Catalunya*, Bellaterra.
- Gracia, C., Burriel J.A., Ibàñez J.J., Mata T., Vayreda J. (2001a) *Inventari Ecològic i Forestal de Catalunya*, Bellaterra.
- Gracia, C., Burriel J.A., Ibàñez J.J., Mata T., Vayreda J. (2001b) *Inventari Ecològic i Forestal de Catalunya*, Bellaterra.
- Graham, R.W. (1992) Late pleistocene faunal changes as a guide to understanding effects of greenhouse warming on the mammalian fauna of North America. In: *Global warming and biological diversity* (eds Peters RL, Lovejoy T), pp. Yale University, New Haven and London.
- Grayson, D.K. (2000) Mammalian response to Middle Holocene climatic change in the Great Basin of the western United States. *Journal of Biogeography* **27**, 181-192.
- Grinnell, J. (1917) The niche-relationships of the California Trasher. *Auk* **34**, 131-135.
- Guisan, A. (2002) A semi-quantitative response model for predicting the spatial distribution of plant species. In: *Predicting Species Occurrences: Issues of Scale and Accuracy* (eds Scott JM, Heglund PJ, Morrison M), pp. Island Press, Covelo, CA.
- Guisan, A., Edwards T.J., Hastie T.J. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* **157**, 89-100.
- Guisan, A., Theurillat J.-P. (2000) Equilibrium modelling of alpine plant distribution: how far can we go? *Phytocoenologia* **30**, 353-384.
- Guisan, A., Theurillat J.-P., Kienast F. (1998) Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science* **9**, 65-74.
- Guisan, A., Weiss S.B., Weiss A.D. (1999) GLM versus CCA spatial modelling of plant species distribution. *Plant Ecology* **143**, 107-122.
- Guisan, A., Zimmermann N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* **135**, 147-186.
- Hagemeijer, W.J.M., Blair M.J. (1997) *The EBCC atlas of European breeding birds, their distribution and abundance*. Poyser, London.
- Harrison, S.P., Prentice I.C. (2003) Climate and CO₂ controls on global vegetation distribution at the last glacial maximum. *Global Ecology & Biogeography* **9**, 983-1004. .
- Hanley, J.A., McNeil B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29-36.
- Hanley, J.A., McNeil B.J. (1983) A method of comparing the areas under receiver operating characteristic (ROC) curve. *Radiology* **148**, 839-843.
- Hastie, T.J., Tibshirani R. (1990) *Generalized additive models*. Chapman and Hall, London.
- Higgins, S.I., Lavorel S., Tackenberg O. (2003) Role of plant spread on plant biodiversity. In: *Climate change and biodiversity: synergistic impacts* (eds Lovejoy TE, Hannah L), pp. Yale University Press.

- Hirzel, A.H., Hausser J., Chessel D., Perrin N. (2002a) Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* **83**, 2027-2036.
- Hirzel, A.H., Hausser J., N. P. (2002b) *Biomapper 2.0*. Division of Conservation Biology, Bern.
- Hirzel, A.H., Helfer V., Métral F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling* **145**, 111-121.
- Hobbs, R.J. (2000) Land-use changes and invasions. In: *The impact of global change on alien species* (eds Mooney HA, Hobbs HA), pp. Island Press, Washington D.C.
- Hosmer, D.W., Lemeshow S. (1989) *Applied Logistic Regression*. Wiley, New York.
- Hu, F.S., Finney B.P., Brubaker L.B. (2001) Effects of holocene Alnus expansion on aquatic productivity, nitrogen cycling, and soil development in southwestern Alaska. *Ecosystems* **4**, 358-368.
- Hughes, L. (2000) Biological consequences of global warming: is the signal already apparent? *TREE* **15**, 56-61.
- Huisman, J., Olff H., Fresco L.F.M. (1993) A hierarchical set of models for species response analysis. *Journal of Vegetation Science* **4**, 37-46.
- Hulme, M., Carter T.R. (2000) The changing climate of Europe. Chapter 3. In: *Assessment of potential effects and adaptations for climate change in Europe: The Europe ACACIA Project* (eds Parry ML), pp. 47-84. Jackson Environment Institute, University of East Anglia, Norwich, UK.
- Huntley, B. (1990) European post-glacial forests: compositional changes in response to climate change. *Journal of Vegetation Science* **1**, 507-518.
- Huntley, B., Bartlein P.J., Prentice I.C. (1989) Climatic control of the distribution and abundance of beech (*Fagus L.*) in Europe and North America. *Journal of Biogeography* **16**, 551-560.
- Huntley, B., Berry P.M., Cramer W., McDonald A.P. (1995) Modelling present and potential future ranges of some European higher plants using climate response. *Journal of Biogeography* **22**, 967-1001.
- Huntley, B., Prentice I.C. (1988) July temperatures in Europe from pollen data, 6000 years before present. *Science* **241**, 687-691.
- Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* **22**, 145-159.
- IPCC (1996) *Climate change 1995: the science of climate change. Contribution of working group I to the second assessment report of the IPCC*. Cambridge University Press, New York.
- Iverson, L.R., Prasad A. (1998) Predicting abundance for 80 tree species following climate change in the Eastern United States. *Ecological Monographs* **68**, 465-485.
- Iverson, L.R., Prasad A., Schwartz M.W. (1999) Modelling potential future individual tree-species distributions in the Eastern United States under climate change scenario: a case study with *Pinus virginiana*. *Ecological Modelling* **115**, 77-93.
- Jalas, J., Suominen J. (1972-1996) *Atlas Florae Europaeae*. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.
- Jansson, R. (2003) Global patterns in endemism explained by past climatic change. *Proceedings of the Royal Society London B* **270**, 583-590.
- Johnson, C.N. (1998) Species extinction and the relationship between distribution and abundance. *Nature* **394**, 272-274.
- Jongman, R.H., ter Braak C.J.F., Van Tongeren O.F.R. (1995) *Data analysis in community and landscape ecology*. Cambridge University Press, Cambridge.

- Kirilenko, A.P., Belotov N.V., Bogatyrev B.G. (2000) Global model of vegetation migration: incorporation of climatic variability. *Ecological Modelling* **132**, 125-133.
- Kleidon, A., Mooney H.A. (2000) A global distribution of biodiversity inferred from climatic constraints: results from a process-based modelling study. *Global Change Biology* **6**, 507-523.
- Koleff, P., Gaston K.J. (2002) The relationship between local and regional species richness and spatial turnover. *Global Ecology & Biogeography* **11**, 363-375.
- Korzukhin, M.D., Ter-mikaelian M.T., Wagner R.G. (1996) Process versus empirical models: which approach for forest ecosystem management. *Canadian Journal of Forest Research* **26**, 879-887.
- Lahti, T., Lampinen R. (1999) From dot maps to bitmaps - Atlas Florae Europaeae goes digital. *Acta Botanica Fennica* **162**, 5-9.
- Lavergne, S., Thuiller W., Molina J., Debussche M. (in prep) Consequences of land use change on rare plant occurrence and extinction in a Mediterranean region: a 115 years study.
- Lavelle, S., Davies I.D., Noble I.R. (2000) LAMOS : a LAndscape MOdelling Shell. In: *Landscape fire modeling-challenges and opportunities. Northern Forestry Centre Information Report NOR-X-371* (eds Hawkes BC, Flannigan MD), pp. 25-28. Natural Resources Canada - Canadian Forest Service.
- Lavelle, S., Garnier E. (2001) Aardvarck to Zyzyxia - functionals groups across kingdoms. *New Phytologist* **149**, 360-363.
- Lavelle, S., Garnier E. (2002) Predicting changes in community composition and ecosystem functioning from plant trait: revisiting the Holy Grail. *Functional Ecology* **16**, 545-556.
- Lavelle, S., Rochette C., Lebreton J.D. (1999) Functional groups for response to disturbance in Mediterranean old field. *Oikos* **84**, 480-498.
- Lawton, J.H. (1994) What do species do in ecosystems? *Oikos* **71**, 367-374.
- Lawton, J.H. (1999) Are there general laws in ecology? *Oikos* **84**, 177-192.
- Leathwick, J.R. (2002) Intra-generic competition among Nothofagus in New Zealand's primary indigenous forests. *Biodiversity and Conservation* **11**, 2117-2187.
- Leathwick, J.R., Austin M.P. (2001) Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* **82**, 2560-2573.
- Leathwick, J.R., Burns B.R., Clarkson B.D. (1998) Environmental correlates of tree alpha-diversity in New Zealand primary forests. *Ecography* **21**, 235-246.
- Legates, D.R., Wilmott C.J. (1990a) Mean seasonal and spatial variability in gauge-corrected, global precipitation. *International Journal of Climatology* **10**, 111-127.
- Legates, D.R., Wilmott C.J. (1990b) Mean seasonal and spatial variability in global surface air temperature. *Theoretical Application Climatology* **41**, 11-21.
- Legendre, P., Legendre L. (1998) *Numerical Ecology*, ELSEVIER edition.
- Lehmann, A., Overton J.M., Austin M.P. (2002) Regression models for spatial prediction: their role for biodiversity and conservation. *Biodiversity and Conservation* **11**, 2085-2092.
- Leibold, M.A. (1995) The niche concept revisited: mechanistic models and community context. *Ecology* **76**, 1371-1382.
- Lek, S., Delacoste M., Baran P., Dimopoulos I., Lauga J., Aulagnier S. (1996) Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling* **90**, 39-52.

- Loehle, C., LeBlanc D. (1996) Model-based assessments of climate change effects on forests: a critical review. *Ecological Modelling* **90**, 1-31.
- Loreau, M. (2000) Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos* **91**, 3-17.
- Mader, H.J. (1990) Wildlife in cultivated landscapes. *Biological Conservation* **54**, 167-173.
- Manel, D., Dias J.M., Buckton S.T., Ormerod S.J. (1999) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology* **36**, 734-747.
- Martín, J., Lopez P. (2002) The effect of Mediterranean dehesa management on lizard distribution and conservation. *Biological Conservation* **108**, 231-219.
- Massolo, A., Meriggi A. (1998) Factors affecting habitat occupancy by wolves in northern Apennines (northern Italy): a model of habitat suitability. *Ecography* **21**, 97-107.
- Maurer, B.A. (2000) Macroecology and consilience. *Global Ecology and Biogeography Letters* **9**, 275-280.
- McCarthy, J.P. (2001) Ecological consequences of recent climate change. *Conservation Biology* **15**, 320-331.
- McCullagh, P., Nelder J.A. (1989) *Generalized linear models*. Chapman & Hall.
- Menzel, A., Estrella N. (2001) Plant phenological changes. In: *"Fingerprints" of climate change - adapted behaviour and shifting species ranges* (eds Walther G-R, Burga CA, Edwards PJ), pp. 123-137. Kluwer Academic Press, New York.
- Midgley, G.F., Hannah L., Millar D., Thuiller W., Booth A. (2003) Developing regional and species-level assessments of climate change impacts on biodiversity: A preliminary study in the Cape Floristic Region. *Biological Conservation* **112**, 87-97.
- Mitchell, T.D. (In prep) A comprehensive set of climate scenarios for Europe.
- Mitchell-Jones, A.J., Amori G., Bogdanowicz W., Krystufek B., Reijnders P.J.H., Spitzenberger F., Stubbe M., Thissen J.M.B., Vohralik V., Zima J. (1999) *Atlas of European Mammals*. Academic Press, London.
- Moisen, G.G., Frescino T.S. (2002) Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling* **157**, 209-225.
- Mooney, H.A., Hobbs H.A. (2000) *The impact of global change on alien species*. Island Press, Washington D.C.
- Mouquet, N., Moore J.L., Loreau M. (2002) Plant species richness and community productivity: why the mechanism that promotes coexistence matters. *Ecology Letters* **5**, 56-65.
- Mourelle, C., Ezcurra E. (1996) Species richness of Argentine cacti: a test of biogeographic hypothesis. *Journal of Vegetation Science* **7**, 667-680.
- Mücher, C.A., editor. (2000) PELCOM project 'development of a consistent methodology to derive land cover information on a European scale from remote sensing for environmental modelling'. Final report Febr. 2000, part 2, EU DG XII-D.
- Nakicenovic, N., Swart R., editors. (2000) Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.
- Ni, J., Sykes M.T., Prentice I.C., Cramer W. (2000) Modelling the vegetation of China using the process-based equilibrium terrestrial biosphere model BIOME3. *Global Ecology and Biogeography* **9**, 463-479.
- Niinemets, N. (2001) Global-scale climatic controls of leaf dry mass per area, density, and thickness in trees and shrubs. *Ecology* **82**, 453-469.
- Nijs, I., Roy J. (2000) How important are species richness, species evenness and interspecific differences to productivity? A mathematical model. *Oikos* **88**, 57-66.

- Oksanen, J., Minchin P.T. (2002) Continuum theory revisited: what shape are species responses along ecological gradients. *Ecological Modelling* **157**, 119-129.
- Olden, J.D., Jackson D.A. (2002) Illuminating the ‘black box’: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling* **154**, 16.
- Openshaw, S., Openshaw C. (1997) *Artificial intelligence in geography*. Wiley.
- Openshaw, S., Turner A. 2000. Forecasting global climatic change impacts on Mediterranean agricultural land use in the 21st Century. Geography Department, University of Leeds, Leeds.
- Ostro, L.E.T., Young T.P., Silver S.C., Koontz F.W. (1999) A geographic information system method for estimating home range size. *Journal of Wildlife Management* **63**, 748-755.
- Parmesan, C., Yohe G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37-42.
- Parry, M.L., editor. (2000) Assessment of potential effects and adaptations for climate change in Europe: The Europe ACACIA Project. Jackson Environment Institute, University of East Anglia, Norwich, UK.
- Pearce, J., Ferrier S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* **133**, 225-245.
- Pearson, R.G., Dawson T.P., Berry P.M., Harrison P.A. (2002) SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecological Modelling* **154**, 289-300.
- Pearson, S.M. (1993) The spatial extent and relative influence of landscape-level factors on wintering bird populations. *Landscape Ecology* **8**, 3-18.
- Peterson, A.T. (2001) Predicting species' geographic distributions based on ecological niche modeling. *The Condor* **103**, 599-605.
- Peterson, A.T. (2003) Projected climate change effects on Rocky Mountain and Great Plain birds: generalities of biodiversity consequences. *Global Change Biology* **9**, 647-655.
- Peterson, A.T., Ortega-Huerta M.A., Bartley J., Sanchez-Cordero V., Soberon J., Buddemeier R.H., Stockwell D.R.B. (2002a) Future projections for Mexican faunas under global climate change scenarios. *Nature* **416**, 626-629.
- Peterson, A.T., Ortega-Huerta M.A., Bartley J., Sánchez-Cordero V., Soberón J., Buddemeier R.H., Stockwell D.R.B. (2002b) Biodiversity consequences of climate change in Mexico.
- Peterson, A.T., Sánchez-Cordero V., Soberón J., Bartley J., Buddemeier R.W., Navarro-Sigüenza A.G. (2001) Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecological Modelling* **144**, 21-30.
- Pigott, C.D., Pigott S. (1993) Water as a determinant of the distribution of trees at the boundary of the Mediterranean zone. *Journal of Ecology* **81**, 557-566.
- Pimm, S.L. (2001) Entrepreneurial insects. *Nature* **411**, 531-532.
- Pimm, S.L., Jones H.L., Diamond J. (1988) On the risk of extinction. *The American Naturalist* **132**, 757-785.
- Pino, J., Rodà F., Ribas J., Pons X. (2000) Landscape structure and bird species richness: implications for conservation in rural areas between natural parks. *Landscape and Urban Planning* **49**, 35-48.
- Poorter, H., Garnier E. (1999) Ecological significance of inherent variation in relative growth rate. In:(eds Valladore FIPaF), pp. 81-103.
- Prentice, I.C. (1986) Vegetation response to past climatic variation. *Vegetatio* **67**, 131-141.

- Prentice, I.C., Bartlein P.J., Webb T. (1991) Vegetation and climate change in eastern North America since the last glacial maximum. *Ecology* **72**, 2038-2056.
- Prentice, I.C., Cramer W., Harrison S.P., Leemans R., Monserud R.A., Solomon A.M. (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography* **19**, 117-134.
- Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters* **3**, 349-361.
- Quézel, P. (1985) Definition of the Mediterranean region and the origin of its flora. In: *Plant conservation in the Mediterranean area* (eds Gomez-Campo C), pp. 9-24. Geobotany, Dordrecht, The Netherlands.
- Rebello, A.G. (1992) *Protea Atlas Manual: instruction booklet to the Protea Atlas Project*. National Botanical Institute, Claremont.
- Rebello, A.G. (2001) *Proteas, a field guide to the Proteas of southern Africa*. Fernwood Press, Vlaeberg, Cape Town.
- Reed, R.A., Peet R.K., Palmer L.W., White P.S. (1993) Scale-dependence on vegetation environment correlation: a case study of a North Carolina piedmont woodland. *Journal of Vegetation Science* **4**, 329-340.
- Reid, P.C., Edwards M., Hunt H.G., Warner A.E. (1998) Phytoplankton change in the North Atlantic. *Nature* **391**, 546.
- Richardson, D.M. (1998) *Ecology and Biogeography of Pinus*. Cambridge University Press.
- Robinson, G.R., Holt R.D., Gaines M.S., Hamburg S.P., Johnson M.L., Fitch H.S., Martinko E.A. (1992) Diverse and contrasting effects of habitat fragmentation. *Science* **257**, 524-526.
- Robinson, R.A., Sutherland W.J. (2002) Post-war changes in arable farming and biodiversity in Great Britain. *Journal of Applied Ecology* **39**, 157-176.
- Robinson, R.A., Wilson J.D., Crick H.Q.P. (2001) The importance of arable habitat for farmland birds in grassland landscapes. *Journal of Applied Ecology* **38**, 1059-1069.
- Root, K.V., Akcakaya H.R., Ginzburg L. (2003a) A Multispecies approach to ecological valuation and conservation. *Conservation Biology* **17**, 196-206.
- Root, T.L., Price J.T., Hall K.R., Schneider S.H., Rosenzweig C., Pounds J.A. (2003b) Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57-60.
- Rouget, M., Richardson D.M., Lavorel S., Vayreda J., Gracia C., Milton S.J. (2001) Determinants of distribution of six *Pinus* species in Catalonia, Spain. *Journal of Vegetation Science* **12**, 491-502.
- Rupp, T.S., Chapin III S.F., Starfield A.M. (2000) Response of subarctic vegetation to transient climatic change on the Seward Peninsula in north-west Alaska. *Global Change Biology* **6**, 541-555.
- Rutherford, M.C., Midgley M.C., Bond W.J., Powrie L.W., Roberts R., Allsopp J. 1999. Plant biodiversity: vulnerability and adaptation assessment. South Africa Country Study on Climate Change.
- Saetersdal, M. (1994) Rarity and species/area relationships of vascular plants in deciduous woods, western Norway - applications to nature reserve selection. *Ecography* **17**, 23-38.
- Saetersdal, M., Birks H.J.B., Peglar S.M. (1998) Predicting changes in Fennoscandian vascular-plant species richness as a result of future climatic change. *Journal of Biogeography* **25**, 111-122.
- Schoener, T.W. (1989) The ecological niche. In: *Ecological Concepts* (eds Cherret J), pp. 790-813. Blackwell.
- Schulze, R.E. 1997. South African Atlas of Agrohydrology and Climatology. Report TT82/96, Water Research Commission, Pretoria.

- Schulze, R.E., Perks L.A. 1999. Assessment of the impact of climate. Final Report to the South African Country Studies Climate Change Programme School of Bioresources Engineering and Environmental Hydrology. University of Natal, Pietermaritzburg, South Africa.
- Schwartz, M.W. (2003) Assessing population viability in long-lived plants. In: *Population viability in plants : conservation, management, and modelling of rare plants* (eds Brigham CA, Schwartz MW), pp. Springer-Verlag, New York.
- Shao, G., Halpin P.N. (1995) Climatic controls of eastern North American coastal tree and shrub distribution. *Journal of Biogeography* **22**, 1083-1089.
- Sitch, S., Smith B., Prentice I.C., Arneth A., Bondeau A., Cramer W., Kaplan J.O., Levis S., Lucht W., Sykes M.T., Thonicke K., Venevsky S. (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ Dynamic Global Vegetation Model. *Global Change Biology* **9**, 161-185.
- Smith, B., Prentice I.C., Sykes M.T. (2001) Representation of vegetation dynamics in modelling of European ecosystems: comparison of two contrasting approaches. *Global Ecology & Biogeography* **10** (ETEMA special issue), 621-637.
- Smith, T.M., Shugart H.H., Woodward F.I. (1997) *Plant Functional Types, Their Relevance to Ecosystem Properties and Global Change*. Cambridge Univ. Press, Cambridge.
- Stephenson, N.L. (1990) Climatic control of vegetation distribution: the role of the water balance. *The American Naturalist* **135**, 649-669.
- Stockwell, D.R.B. (1999) Genetic algorithms II. In: *Machine Learning Methods for Ecological Applications* (eds Fielding AH), pp. Kluwer Academic Publishers.
- Stockwell, D.R.B., Arzberger P., Fountain T., Helly J. (2000) An interface between computing, ecology and biodiversity: environmental informatics. *Korean J. Ecol* **23**, 101-106.
- Stockwell, D.R.B., Peters D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *Int. J. Geographical Information Science* **13**, 143-158.
- Swets, K.A. (1988) Measuring the accuracy of diagnostic systems. *Science* **240**, 1285-1293.
- Swihart, R.K., Gehring T.M., Kolozsvary M.B. (2003) Responses of 'resistant' vertebrates to habitat loss and fragmentation: the importance of niche breadth and range boundary. *Diversity and Distribution* **9**, 1-18.
- Sykes, M.T., Prentice I.C., Smith B., Cramer W., Venevsky S. (2001) An introduction to the European ecosystem modelling activity. *Global Ecology and Biogeography* **10**, 581-593.
- Taberlet, P., Cheddadi R. (2002) Quaternary refugia and persistence of biodiversity. *Science* **297**, 2009-2010.
- Taberlet, P., Fumagali L., Wust-Saucy A.G., Cosson J.F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **6**, 289-301.
- ter Braak, C.J.F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**, 1167-1179.
- ter Braak, C.J.F., Prentice I.C. (1987) A theory of gradient analysis. *Chapitre 9. Thèse de ter Braak (1987)*.
- Thuiller, W. (2003) BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* **9**, 1353-1362.
- Thuiller, W., Araújo M.B., Lavorel S. (2003a) Generalized Models versus Classification Tree Analysis: a comparative study for predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* **14**, 669-680.

- Thuiller, W., Araújo M.B., Lavorel S. (in press-a) Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography*.
- Thuiller, W., Brotons L., Araújo M.B., Lavorel S. (in press-b) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*.
- Thuiller, W., Lavorel S., Midgley G.F., Lavergne S., Rebelo A.G. (in press-c) Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* species in the Cape Floristic Region. *Ecology*.
- Thuiller, W., Vaydera J., Pino J., Sabaté S., Lavorel S., Gracia C. (2003b) Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology & Biogeography* **12**, 313-325.
- Tilman, D., Lehman C.L., Thomson K.T. (1997) Plant diversity and ecosystem productivity: theoretical considerations. *Proc. Natl. Acad. Sci. USA* **94**, 1857-1861.
- Tilman, D., Pacala S. (1993) The maintenance of species richness in plant communities. In: *Species diversity in ecological communities* (eds R.E. RaDS), pp. 13-25. University Chicago Press, Chicago.
- Tutin, T.G., Burges N.A., Chater A.O., Edmondson J.R., Heywood V.H., Moore D.M., Valentine D.H., Walters S.M., Webb D.A. (1964-1993) *Flora Europaea*. Cambridge University Press, Cambridge.
- Tyson, P., Odada E., Schulze R., Vogel C. (2002) Regional-global change linkages: South Africa. In: *Global-regional linkages in the earth system* (eds Tyson P, Fuchs R, Fu C, et al.), pp. 3-73. Berlin, Springer.
- van Humboldt, A., Bonpland A. (1807) *Essai sur la géographie des plantes*, Paris.
- Vayssières, M.P., Plant R.E., Allen-Diaz B.H. (2000) Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* **11**, 679-694.
- Venables, W.N., Ripley B.D. (2002) *Modern applied statistic with S*. Springer.
- Visser, M.E., Holleman L.J.M. (2001) Warmer springs disrupt the synchrony of oak and winter moth phenology. *Proceedings of the Royal Society London B* **268**, 289-294.
- Visser, M.E., Van Noordwijk A.J., Tinbergen J.M., Lessells C.M. (1998) Warmer springs lead to mis-times reproduction in great tits (*Parus major*). *Proceedings of the Royal Society London B* **265**, 1867-1870.
- Vitousek, P.M., D'Antonio C., Loope L.L., Rejmanek M., Westbrooks R. (1997) Introduced species: a significant component of human-caused global change. *N.Z. J. of Ecology* **21**, 1-16.
- Vitousek, P.M., Walker L.R. (1989) Biological invasion by *Myrica faya* in Hawaii: plant demography, nitrogen fixation, ecosystem effects. *Ecological Monographs* **59**, 247-265.
- Walther, G.-R., Post E., Convey P., Menzel A., Parmesan C., Beebee T.J.C., Fromentin J.-M., Hoegh-Gudberg O., Bairlein F. (2002) Ecological responses to recent climate change. *Nature* **416**, 389-395.
- Warren, M.S., Hill J.K., Thomas J.A., Asher J., Fox R., Huntley B., Roy D.B., Telfer M.G., Jeffcoate S., Harding P., Jeffcoate G., Willis S.G., Greatorex-Davies J.N., Moss D., Thomas C.D. (2001) Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature* **414**, 65-69.
- Watson, R.T., et. a.l., editors. (2001) IPCC Third assessment report: Climate change 2001. Synthesis Report. IPCC, Geneva, Switzerland.
- Webb, T.I. (1981) The past 11,000 years of vegetational change in eastern North America. *BioScience* **31**, 501-506.
- Webb, T.I. (1992) Past changes in vegetation and climate: lessons for the future. In: *Global warming and biological diversity* (eds Peters RL, Lovejoy TE), pp. Yale University, New Haven and London.

- Whisenant, S.G. (1991) Changing fire frequencies on Idaho's Snake River plains: Ecological and management implications. In: *Proceedings-Symposium of cheatgrass invasion, shrub die-off, and other aspects of shrub biology and management* (eds McArthur ED, Rommey EM, Smith SD, et al.), pp. 4-10. Intermountain Research Station, Forest Service, USDA, Ogden.
- Willis, K.J., Whittaker R.J. (2002) Species diversity-scale matters. *Science* **295**, 1245-1248.
- Woodward, F.I. (1987) *Climate and plant distribution*. Cambridge University Press, Cambridge.
- Woodward, F.I., Cramer W. (1996) Plant functional types and climate changes: Introduction. *Journal of Vegetation Science* **7**, 306-308.
- Woodward, F.I., Williams B.G. (1987) Climate and plant distribution at global and local scales. *Vegetatio* **69**, 189-197.
- Yee, T.W., Mitchell N.D. (1991) Generalized additive models in plant ecology. *Journal of Vegetation Science* **2**, 587-602.
- Zaniewski, E.A., Lehmann A., Overton J. (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., Kienast F. (1999) Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science* **10**, 469-482.

ANNEXE 1.1

Generalized models vs classification tree analysis: predicting spatial distributions of plant species at different scales

Thuiller W., Araújo M.B., Lavorel S.

Journal of Vegetation Science (2003) **14**, 669-680

Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales

Thuiller, Wilfried*; Araújo, Miguel B. & Lavorel, Sandra

Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de Recherche Scientifique, 1919 route de Mende,
F-34293 Montpellier Cedex 5, France; *Corresponding author; E-mail wilfried.thuiller@cefe.cnrs-mop.fr

Abstract. Statistical models of the realized niche of species are increasingly used, but systematic comparisons of alternative methods are still limited. In particular, only few studies have explored the effect of scale in model outputs. In this paper, we investigate the predictive ability of three statistical methods (generalized linear models, generalized additive models and classification tree analysis) using species distribution data at three scales: fine (Catalonia), intermediate (Portugal) and coarse (Europe). Four Mediterranean tree species were modelled for comparison. Variables selected by models were relatively consistent across scales and the predictive accuracy of models varied only slightly. However, there were slight differences in the performance of methods. Classification tree analysis had a lower accuracy than the generalized methods, especially at finer scales. The performance of generalized linear models also increased with scale. At the fine scale GLM with linear terms showed better accuracy than GLM with quadratic and polynomial terms. This is probably because distributions at finer scales represent a linear sub-sample of entire realized niches of species. In contrast to GLM, the performance of GAM was constant across scales being more data-oriented. The predictive accuracy of GAM was always at least equal to other techniques, suggesting that this modelling approach is more robust to variations of scale because it can deal with any response shape.

Keywords: Accuracy assessment; Climate variable; Generalized additive model; Generalized linear model; Mediterranean tree species; Receiver operating characteristics curve.

Abbreviations: GLM = Generalized Linear Model; GAM = Generalized Additive Model; CTA = Classification Tree Analysis; ROC curve = Receiver Operating Characteristic curve; AUC = Area Under the Curve.

Introduction

Modelling plant distributions in relation to environmental factors has gained momentum with recent developments in the fields of GIS and statistical techniques applied to ecological analysis (e.g. Guisan & Theurillat 2000). When they include climate data, these models can be used to make projections of species' distributional changes in response to climate change (Bakkenes et al. 2002; Huntley et al. 1995; Midgley et al. 2003).

One of the common approaches in plant bioclimatic modelling is the statistical fit of correlative models, where plant responses are fitted by regression to some environmental variables. Some of these methods are based on the idea of bell-shaped species responses along environmental gradients as used in regression and ordination-based vegetation response models (e.g. Austin 1985; ter Braak 1987). However, several authors (e.g. Austin & Gaywood 1994; Huisman et al. 1993) have shown that asymmetric and other complex response curves are very frequent. Methods to overcome this problem include generalized linear models (GLM), with skewed response curves fitted by a β -function or third-order cubic polynomial terms (Austin et al. 1994), generalized additive models (GAMs) (Yee & Mitchell 1991), regression and classification tree analyses (RTA and CTA) (Breiman et al. 1984) and other rule-based methods such as neural networks (Olden & Jackson 2002; Pearson et al. 2002) or cellular automata (Carey 1996), for which no assumptions are made regarding species response curves to environmental gradients.

There are few studies comparing the predictive accuracy of statistical and rule-based methods (but see Bio et al. 1998; Franklin 1998; Pearce & Ferrier 2000b; Vayssières et al. 2000; Vetaas 2000; Walker & Cocks 1991). Guisan et al. (1999) compared the predictive accuracy of GLM and canonical correspondence analysis (CCA), and showed that GLM provided better models for single species while CCA provided a broader overview of multiple species assemblages. These results provide an interesting but still limited exploration of which method would perform best for different goals

and types of data. One of the problems with these analyses is that results are very much dependent upon the mathematical procedure used. There are often several procedures to fit the same model.

Model results are also dependent on the adequacy and quality of environmental data. To provide good predictions of current species distribution, caution must be given to the selection of environmental variables for analysis. By order of preference, it is advisable to select fewer relevant climatic variables that are suitable surrogates of direct physiological parameters of plants (Woodward 1987). In addition to these, soil type, slope and elevation are often considered as good indirect variables to predict plant distributions, even if their link with physiological mechanisms is more complex (Burke 2001; Nicholls 1989). The relative importance of environmental factors to determine plant species distributions also varies according to spatial scale (Huston 1994). At the community and landscape scales direct physiological parameters, species dispersal abilities, biotic interactions and local disturbances control plant distributions. At continental and global scales, macroclimate is the major determinant of plant species distributions (Woodward 1987, 1990). At intermediate scales, human impact or management can be particularly important, but few studies have included such variables because of the paucity of adequate digital data sets. Indeed, many regression models have been fitted to site data with a variety of predictors, including climate, land use or intensity of disturbance, but without concern for regional or global prediction. When projections to other sites are intended, complete coverage of the predictors for the region is required but rarely available. Finally, at any scale, plant distribution models are mainly dependent on the quality, adequacy and resolution of the environmental data.

The aim of this study was to compare the predictive accuracy of one familiar parametric approach, generalized linear models (GLM), and two non-parametric approaches that have recently gained popularity in ecology, generalized additive models (GAM) and classification tree analysis (CTA). The comparison was performed using three independent data sets representing three scales and resolutions: fine (Catalonia), intermediate (Portugal) and coarse (Europe). Bioclimatic models were calculated for four common tree species and a range of environmental variables to address the following questions: 1. How do the different methods respond to increases in scale? 2. Which method is more adequate to fine, intermediate or coarse scales? 3. Which are the best environmental variables for a given scale? 4. What methodological framework might be more robust and accurate to generate projections of current plant distributions in Europe?

Data and Methods

Data sets

Study species

We selected four Mediterranean species for analysis: *Juniperus oxycedrus*, *Pinus pinaster*, *Pinus pinea* and *Quercus suber*. The choice of common species took into consideration the number of records per species in our databases, as both very rare and very common species are bound to be more difficult to model.

J. oxycedrus is a dioecious tree reaching 14 m, but is usually encountered as a shrub of oval habit. It occurs between 0 and 1400 m a.s.l. and is also found in more continental locations of Europe and W Asia.

P. pinaster is a larger single-stemmed tree (30-40 m), mainly found in the W Mediterranean. In continental Europe it occurs from 0 to 400 m a.s.l., but reaches up to 900 m in Corsica and to 2000 m in Morocco.

P. pinea is a shorter pine tree (12-25 m) often identified with Mediterranean landscapes. It is probably original to the Iberian Peninsula, but an archaeophyte throughout the Mediterranean region (Barbéro et al. 1998).

Q. suber occurs all over the Mediterranean in open woodlands on hills and lower slopes, generally on acidic soils. It is grown commercially for its thick cork bark for which Portugal is the world's largest producer.

Fine-scale data from Catalonia

Species data: These are from a database of the Forestry Inventory of Catalonia (IEFC) managed by the Centre for Ecological Research and Applied Forestry in Barcelona (CREAF). Sampling areas have a circular shape usually with a radius of 10 m.

Environmental data: A climate model (Ninyerola et al. 2000) was used to generate layers of monthly precipitation and monthly mean, minimum and maximum temperatures for the entire study area (Table 1). This model used a network of weather stations in Catalonia (257 stations for precipitation and 160 stations for temperatures). From these data, a multiple regression analysis was performed between the meteorological variables and a set of geographic variables (latitude, solar radiation, altitude, distance to the coast and cloudiness factors) derived from a 180-m resolution Digital Elevation Model (DEM). Accuracy of the climate surfaces was assessed and corrected with reference to an independent set of weather stations (40% of the initial weather stations). Layers of climatic variables (Table 1) were then derived from the original climate model. Topographic layers (elevation, slope and solar radiation) were provided by a DEM (180 m resolution). A geology layer was provided by the Institute of Cartography of Catalonia (1 : 250000).

The data were used in raster format in ArcView (Vers. 3.2a for Windows, ESRI Corp.) with a resolution of 180 m. We combined raster environmental data and vegetation point data to determine the value of each environmental variable in each vegetation plot (Table 1).

Intermediate-scale data from Portugal

Species data: Data were compiled from an ongoing database project at the University of Évora (www.cea.uevora.pt/umc). Presence-absence data were referenced onto 993 UTM 10 km × 10 km grid cells.

Environmental data: Climatic and geomorphologic data (Table 1) were compiled from available digital layers of the ‘Atlas do Ambiente’ (http://www.dga.pt). Data were converted from available vector maps and re-sampled to 10 km × 10 km grid maps, using zonal functions in ArcView to calculate the mode of each variable in each cell (Table 1).

Coarse scale from Europe

Species data: Species presence data were a subset of the Atlas Florae Europaeae (AFE) (Jalas & Suominen 1972-1996) digitized by Lahti & Lampinen (1999). Data are referenced onto 4419 UTM 50 km × 50 km grid cells. This grid offers mapping units across Europe.

Environmental data: Seven environmental variables were selected and converted from 0.5° latitude-longitude maps to UTM grid cells (Table 1): mean annual precipitation (Legates & Wilmott 1990a); mean temperature in January and July (Legates & Wilmott 1990b); mean annual potential evapotranspiration and mean potential evapotranspiration in January and July (Ahn & Tateishi 1994) and altitude, obtained from the United Nations Environmental Program grid. Temperature and precipitation were extracted from the National Climatic Data Center (NOAA). Data were re-sampled with the GIS package IDRISI (Eastman 1996).

Comparability of environmental data

Environmental variables used for analysis are not entirely comparable across the three scales (Table 1). However, they can be grouped according to shared impacts on plant growth, development and potential distribution. Following Austin & Smith (1989), we distinguished two types of variables. First, direct variables that have direct physiological impacts on plant growth. Examples include temperature, precipitation or radiation. The three variables were available in the Catalanian environmental data set, while only precipitation and temperature were available within the Portuguese and European environmental data sets. In addition, the Catalanian data contained not only mean annual precipitation (MAP), like the others, but also seasonal

Table 1. Environmental variables used to construct the different models. prec. = precipitation; temp. = temperature; evap. = evapotranspiration; SD = standard deviation. Cat. = Catalonia; Port. = Portugal; Eur. = Europe.

		Cat.	Port.	Eur.
Mean annual prec.	MAP	X	X	X
Mean summer prec.	SumP	X		
Mean winter prec.	WinP	X		
Mean Autumn prec.	AutP	X		
Mean Spring prec.	SprP	X		
Mean Annual temp.	MAT	X	X	
Mean temp. coldest month	MTC	X		X
Mean temp. warmest month	MTW	X		X
Mean annual max. temp.	MAMxT	X		
Mean max temp. hottest month	MMTW	X		
Mean annual min. temp.	MAMnT	X		
Mean min. temp. coldest month	MMTC	X		
Mean annual radiation	MAR	X		
Annual potential evap.	PET			X
Potential evap. coldest month	PETC			X
Potential evap. warmest month	PETW			X
Elevation	Elev	X	X	X
Max. elevation	MxElev		X	
Min. elevation	MnElev		X	
SD elevation	SdElev		X	
Slope	Slope	X	X	
SD slope	SdSlope		X	
Variation slope coefficient	VCoSlope		X	
Geology	Geol	X		

precipitation. Mean of the coldest and warmest months (MTC and MTW, respectively) were available in the Catalanian and European data, but not in the Portuguese data, which provided mean annual temperature (MAT) instead. This little difference, in which direct gradients were quantified for each data set made it possible to compare the significance of variables that were either common to all data sets (seasonal precipitation, MTC, MTW) or were exclusive to some of them (MAP, MAT). Second, indirect variables are those such as topography and geology, both of which were represented in our three environmental databases. Environmental data from Catalonia contained Mean Elevation, Slope and Geology. Portuguese data contained four variables related to elevation (Table 1), three variables related to slope (Table 1) and geology (Table 1), while Europe contained only Mean Elevation. Hence the three environmental data contained similar variables that could be related to each other for comparisons.

Statistical models

Statistical analyses were performed using Splus (Vers. 5 for Windows, MathSoft Inc.), with standard functions (*gam* for generalized additive models, *glm* for generalized linear models and *tree* for classification tree analysis) and some custom functions. Species distribution maps were created with ArcView and its Spatial

Analyst extension (Vers. 3.2a for Windows, ESRI Corp.). In order to evaluate the quality of predictions, we divided databases into two subsets: calibration and evaluation. The first, a random sample from 70% of total database, was used to calibrate the models, whereas the second, comprising of the remaining data, was used to evaluate model predictions (Fielding & Bell 1997).

Generalized Linear Models

GLMs are often used to describe the relationship between species and their environment (see Guisan & Zimmermann 2000). GLMs provide a less restrictive form than classic multiple regressions by providing error distributions for the dependent variable other than normal and non-constant variance functions (McCullagh & Nelder 1989). If the response with a predictor variable is not linear, then a transformation can be included; polynomial terms are allowed for the simulation of skewed and bimodal responses (Guisan et al. 1999), β -functions (Austin & Gaywood 1994) or a hierarchical set of models (Huisman et al. 1993). The nature of the relationship between species and environmental gradients has to be known *a priori*.

We developed two kinds of GLM. First, we examined the relationship between the species presence-absence response and environmental predictors. We then performed a simple GLM with linear terms allowing for the possibility of interactions. A complex GLM was performed with the same parameters, to which were added second and third order terms and polynomial terms (second and third orders). In this case, we did not consider interactions between quadratic and polynomial terms because their biological significance would be difficult to interpret.

To select for the most parsimonious model, we used an automatic stepwise model selection using the Akaike Information Criterion (AIC) (Chambers & Hastie 1997). This procedure allowed the removal of redundancy in variables and eliminated multicollinearity problems, even if principal correspondence analysis on environmental data sets did not show strong patterns of correlations.

Generalized Additive Models

GAMs are designed to capitalize on the strengths of GLMs without requiring the problematic steps of postulating a response curve shape or specific parametric response function. They use a class of equations called 'smoothers' that attempt to generalize data into smooth curves by local fitting to subsections of the data. GAMs are therefore useful when the relationship between the variables is expected to be of a more complex form, not easily fitted by standard linear or non-linear models, or where there is no *a priori* reason for using a particular model (Hastie & Tibshirani 1990).

We used a cubic spline smoother, which is a collection of polynomials of degree less than or equal to 3, defined on subintervals. A separate polynomial is fitted for each neighbourhood, thus enabling the fitted curve to join all of the points. The degree of smoothness was automatically selected by cross-validation and restricted to 4. Similarly to GLM, we used an automated stepwise process (the *step.gam* function in Splus) to select the most significant variables for each species.

Regression and Classification Trees

These provide an alternative to regression techniques (e.g. Vayssières et al. 2000; Thuiller et al. 2003). They do not rely on *a priori* hypotheses about the relation between independent and dependent variables. This method consists of recursive partitions of the dimensional space defined by the predictors into groups that are as homogeneous as possible in terms of response. The tree is built by repeatedly splitting the data, defined by a simple rule based on a single explanatory variable. At each split, the data are partitioned into two exclusive groups, each of which is as homogeneous as possible (Thuiller 2003). To control the length of the tree, we use the *prune* function of Splus (Vers. 5 for Windows, MathSoft Inc.). The program builds a nested sequence of subtrees by recursively snipping off the less important splits in terms of explained deviance. The length of the tree was controlled by choosing the best trade-off between explained deviance and tree size. Each predictor could be used several times in the tree if it improved the predictive performance.

Predicting assessment accuracy

To compare different models from different species using records from different databases, we were reluctant to use an accuracy index dependent on any fixed threshold. Instead we used the Receiver Operating Characteristic curve (ROC curve) that is not dependent on the threshold. The use of the ROC curve method is still in its infancy in ecology (Pearce & Ferrier 2000b; Thuiller et al. 2003) as compared to the more widely used κ or Classification Accuracy (e.g. Franklin 1998; Guisan et al. 1998; Huntley 1995). The ROC curve is a graphical method that represents the relation between the False Positive fraction (1 – specificity) and the sensitivity for a range of thresholds. If all predictions were possibly expected by chance, the relation would be a 45° line. Good model performance is characterized by a curve that maximizes sensitivity for low values of (1-specificity), i.e. when the curve passes close to the upper left corner of the plot. The area between the 45° line and the curve measures discrimination, that is, the ability of the model to correctly classify a species as present or absent in a given plot (area under the curve: AUC). A rough

guide for classifying the accuracy of a diagnostic test is the traditional academic point system (Swets 1988): 0.90-1.00 = excellent; 0.80-0.90 = good; 0.70-0.80 = fair; 0.60-0.70 = poor; 0.50-0.60 = fail.

The difference between the areas under ROC curves generated by two or more models provides a measure of comparative discrimination ability of these models when applied to independent evaluation data. In our study, the significance of the difference between two ROC curves (AUC1 and AUC2) can be calculated as a critical ratio test (Hanley & McNeil 1983):

$$Z = \frac{AUC1 \pm AUC2}{\sqrt{Se_{AUC1} + Se_{AUC2} * Se_{AUC1} Se_{AUC2}}} \quad (1)$$

Se is the standard error and r is the correlation between both areas under the curve. r is the mean of the correlation between predictions from both models for the positive events and the correlation between predictions for the negative events, calculated using the Spearman rank correlation coefficient. To project species distributions into binary presence-absence form, we used a probability threshold maximising the percentage of presence and absence correctly predicted.

To address our questions we have articulated the results in two parts, within-scale and across-scale.

Results

Within-scale comparisons

Fine-scale Catalonia

At this scale all models provided generally good results for all species (Table 2). The highest model accuracies were observed for *P. pinea* and *Q. suber* (AUC > 0.9), while slightly lower but still high accuracies were obtained for *J. oxycedrus* and *P. pinaster* (AUC > 0.82 and AUC > 0.85 respectively). *Q. suber* was the species best predicted, with on average 94% of occurrences correctly predicted, while for *J. oxycedrus* only 76% of the occurrences were correctly predicted. For *J. oxycedrus* and *P. pinaster* GLM 'simple' and GAM had significantly higher discrimination ability ($p < 0.05$) than GLM 'complex' and CTA whereas for *P. pinea* and *Q. suber* all types of general models (GLM 'simple', 'complex' and GAM) were significantly better than CTA ($p < 0.05$).

The most significant environmental variables were related to winter climate (coldest month mean temperature and winter precipitation) for *Pinus* and *Q. suber*, but for *J. oxycedrus* the most important variables were related to summer and annual precipitation.

Some differences were found between the discrimi-

nation abilities of species. The distribution of *Q. suber* was well simulated, but *J. oxycedrus* and *P. pinaster* were poorly predicted. As an illustration, we plotted the predicted species distribution of *Q. suber* according to GAM (one of the highest AUC values) and CTA (the lowest AUC value) (Fig. 1). Both predicted distributions were similar, slightly over-predicting the distribution of *Q. suber* in the south. The northern distribution margin was better predicted by GAM than by CTA. For example, CTA predicts two presences in northern Catalonia, where *Q. suber* has not been recorded.

A trend of accuracy of models is discernible from the analyses. CTA models had significantly lower discrimination ability than generalized models at this scale. GLM 'simple' and GLM 'complex' had similar results showing that more complex models do not always perform better than simple ones. Ease of interpretation and parsimony would therefore make GLM 'simple' the best model choice at the fine scale.

Intermediate-scale Portugal

For this data set and scale there were strong disparities between models for individual species and these were not consistently related to the type of method used (Table 3). *J. oxycedrus* was the best-predicted species (mean AUC = 0.90), whereas models for *Q. suber* were not accurate (mean AUC = 0.635). Both species of *Pinus* had a fairly accurate prediction ($0.741 < AUC < 0.804$) across all models. There were no significant differences among models for any species except for *P. pinea*, where GLM 'simple' had a better accuracy than GLM 'complex' and GAM, and for *P. pinaster* where GAM had a better accuracy than GLM 'complex'.

Relevant environmental variables for *J. oxycedrus* and *P. pinaster* were related to annual precipitation, while elevation explained most of the variation for *P. pinea*. For *Q. suber*, different models tended to select inconsistent sets of variables.

We selected the distributions predicted by GLM 'complex' and GAM. *J. oxycedrus* has an aggregated distribution and is restricted to NE Portugal (Fig. 2). Both GLM 'complex' and GAM over-predicted the observed distribution of this species. The major difference between these two models was that GAM had a lower rate of false positive, over-predicting less than GLM 'complex'. The fact that both models over-predicted the distribution of *J. oxycedrus* in the same areas could reflect the potential distribution of the species, which may not yet have colonized this part of Portugal or may be restricted to the northeast by human activities.

Hence there was no general trend in accuracy among models, and no model seemed to be better than others at this scale.

Species	Model	Selected environmental variables	Evaluation		
			se	sp	AUC
<i>Juniperus oxycedrus</i>	GLMs	MAP, SdSlope, Geol, MAP:SdSlope	93.3	92.8	0.939
	GLMs	AugP, Geol, MAP, MAP:SumP	76.3	76.3	0.871 *
	GLMc	pol(AugP,3), pol(MAP,2), Geol, SumP^3	76.3	76.3	0.859
	GAM	s(AugP), s(WinP), s(MAMxT), s(SumP)	79.0	78.3	0.865 *
<i>Pinus pinaster</i>	CTA	AugP, MTC, MMTW	65.8	86.3	0.813
	GLMs	MTC, SprP, WinP, Elev	82.3	82.2	0.869 *
	GLMc	pol(MTC,2), WinP^3, pol(SumP,2), pol(MMTC,3)	82.3	81.9	0.862
<i>Pinus pinea</i>	GAM	s(MTC), s(WinP), s(MAMxT), s(SprP)	82.3	82.6	0.866 *
	CTA	MTC, WinP, Elev, SumP	83.5	77.1	0.816
	GLMs	MTC, SumP:AugP, MAP:SumP, AugP	83.3	83.2	0.904 *
	GLMc	MTC^3, pol(SumP,3), MAMxT^3, pol(MMTC,3)	84.1	84.0	0.911 *
<i>Quercus suber</i>	GAM	s(MTC), Geol, s(MAMnT), s(WinP)	84.0	84.0	0.911 *
	CTA	MTC, WinP, MAMxT, MAP	87.1	77.1	0.880
	GLMs	Geol, MTC, WinP, WinP:MTC	95.4	95.5	0.988 **
	GLMc	pol(MTC,3), Geol, pol(WinP,3), MAMnT	94.5	94.6	0.986 **
Mean over species	GAM	s(MTC), Geol, s(MAMnT), s(WinP), CTA	94.1	94	0.985 *
	CTA	MTC, Geol, WinP, MMTW	91.6	93.6	0.952
	GLMs		84.3	84.3	0.908
	GLMc		84.3	84.2	0.905
	GAM		84.9	84.7	0.907
	CTA		82.0	83.5	0.865

Table 2. Prediction accuracy on evaluation data of the different models for each species in the Catalonia data set: The four most important selected environmental variables influencing are presented in order of decreasing deviance explained. Se = sensitivity; sp = specificity (%). * = model significantly better than other models at $p = 0.05$; ** $p = 0.01$. GLMc = GLM ‘complex’; GLMs = GLM ‘simple’. E.g. pol(MAP,2) generate a matrix of orthonormal polynomials; (MAP+MAP^2). SumP^3 = variable summer precipitation at third degree; MAP:SumP = interaction between mean annual and summer precipitation.

Species	Model	Selected environmental variables	Evaluation		
			se	sp	AUC
<i>Juniperus oxycedrus</i>	GLMs	MAP, SdSlope, Geol, MAP:SdSlope	93.3	92.8	0.939
	GLMc	pol(MAP,2), SdSlope^2, pol(MnElev,2), pol(MxElev,3)	80	80	0.929
	GAM	s(SdSlope), s(MAP), s(Elev), s(MAT)	86.7	86.7	0.941
	CTA	MxElev, MAP, Slope, SdElev	80	90	0.936
<i>Pinus pinaster</i>	GLMs	MAP, MAP:SdSlope, Geol, SdSlope	70.3	70.5	0.787
	GLMc	pol(MAP,3), pol(VCoElev,3), Geol, pol(SdSlope,2)	71	71.2	0.782
	GAM	s(MAP), s(VcoElev), s(Elev), Geol	73.9	73.7	0.804 *
	CTA	MAP, VCoElev, Elev, MAT	73.9	71.2	0.785
<i>Pinus pinea</i>	GLMs	Geol, Elev, SdElev, MAP	70.8	70.3	0.766 **
	GLMc	pol(Elev,3), Geol, SdElev, MAP	70.8	70.7	0.745
	GAM	s(Elev), Geol, s(SdElev), s(MAP)	70.8	70.3	0.748
	CTA	Elev, Sd.Slope, MAT, SdElev	73.6	70.7	0.741
<i>Quercus suber</i>	GLMs	SdSlope, Elev, MAP:SdSlope, MxElev	60.3	60.2	0.630
	GLMc	pol(MAP,3), pol(Elev,2), pol(MnElev,2), Elev^2	60.3	60.2	0.651
	GAM	s(Elev), s(MAP), s(MnElev), s(VcoElev)	57.4	58	0.635
	CTA	VCoElev, Elev, Geol, Slope	47.1	67.3	0.628
Mean over species	GLMs		73.7	73.5	0.781
	GLMc		70.5	70.5	0.777
	GAM		72.2	72.2	0.782
	CTA		68.7	74.8	0.773

Table 3. Prediction accuracy on evaluation data of the different models for each species in the Portugal data set: The four most important selected environmental variables influencing are presented in order of decreasing deviance explained. See further Table 2.

Species	Model	Selected environmental variables	Evaluation		
			se	sp	AUC
<i>Juniperus oxycedrus</i>	GLMs	PET, MTW:PET, PETC, Elev	87	87.2	0.936
	GLMc	pol(PET,3), pol(MTW,3), pol(Elev,3), pol(PETC,2)	88	88.1	0.941
	GAM	s(MTW), s(Elev), s(MTC), s(PETW)	88	88.1	0.937
	CTA	PET, PETW, MTW, Elev	87	86.8	0.927
<i>Pinus pinaster</i>	GLMs	MTC, Elev, PETC, MTC:MTW	87.2	87.8	0.935
	GLMc	pol(PET,3), pol(MTW,3), pol(MTC,3), pol(PET,3)	85.1	86	0.947
	GAM	s(MTC), s(Elev), s(MTW), s(PET)	87.2	86.5	0.952 *
	CTA	PET, MTC, MTW, Elev	87.2	87.8	0.926
<i>Pinus pinea</i>	GLMs	PET, MTC, MTC:PET, MTC:PETC	82.4	82.8	0.901
	GLMc	pol(PET,3), MTW^3, pol(MTW,2), pol(MTC,3)	85.3	85.4	0.931 *
	GAM	s(MTC), s(PETW), s(PETC)	82.4	83.6	0.923 *
	CTA	PET, MTC, MTW, PETW	85.3	84.7	0.879
<i>Quercus suber</i>	GLMs	MTC, PET, PETC, MTC:PET	88.7	89	0.941
	GLMc	pol(PET,3), pol(MTC), pol(MTW), pol(Elev,3)	90.6	90.6	0.960 *
	GAM	s(MTC), s(MTW), s(PET)	90.6	90.6	0.962 *
	CTA	PET, MTC, MTW, Elev	92.5	87.9	0.945
Mean over species	GLMs		86.3	86.7	0.928
	GLMc		87.3	87.5	0.945
	GAM		87.1	87.2	0.944
	CTA		88.0	86.8	0.919

Table 4. Prediction accuracy on evaluation data of the different models for each species in the European data set: The four most important selected environmental variables influencing are presented in order of decreasing deviance explained. See further Table 2

Coarse-scale Europe

Models at the European scale had generally high AUC values (Table 4). For *Q. suber* they were very high ($0.945 < \text{AUC} < 0.962$), and GLM 'complex' and GAM were significantly better than CTA and GLM 'simple'. The same pattern was found for *P. pinea* ($0.879 < \text{AUC} < 0.931$). GAM was also better than GLM 'simple' for *P. pinaster* ($0.926 < \text{AUC} < 0.952$), but there were no significant differences between models for *J. oxycedrus* ($0.927 < \text{AUC} < 0.936$). The two variables most frequently selected by the models were mean temperature of the coldest month and annual potential evapotranspiration. Elevation was important for *J. oxycedrus* and *P. pinaster*, whereas mean temperature of the warmest month was selected in second or third position for *P. pinea* and *Q. suber*.

Though differences between models tended to be small at this scale, two groups could be distinguished: GAM and GLM 'complex' had a better discrimination ability and accuracy ($\text{AUC} = 0.944$ and $\text{AUC} = 0.945$ respectively) than GLM 'simple' and CTA. The predictive ability of different types of models was similar across species although these covered a wide range of occurrence values (*J. oxycedrus*: 326; *P. pinaster*: 157; *P. pinea*: 114; *Q. suber*: 183 occurrence records on 2209 available grid cells) and had rather different geographical distributions. We plotted the predicted distribution of *Q. suber* by GLM 'simple' and GAM models (Fig. 3). Both models over-predicted its actual distribution in Spain. However, the errors generated by the two models showed different spatial patterns. GLM 'simple' over-predicted species distribution in Greece while GAM was closer to reality. Inversely, GAM over-predicted the distribution in southwestern France while GLM reproduced almost perfectly the observed distribution.

Across-scale comparisons

Models

A comparison of models across scales revealed some distinctive patterns in model performance among species. At the finest scale GLM 'simple' and GAM provided better discrimination ability than GLM 'complex' and CTA for all species. In contrast, at the intermediate scale there was no best model (except GLM 'simple' for *P. pinea*). At the largest scale GLM 'complex' and GAM discriminated better and GLM 'simple' seemed to be the least accurate. Hence, differences in performance between GLM with linear terms (and interaction) and GLM with linear (and interaction), quadratic and polynomial terms appeared to be scale-dependent. At the fine scale GLM 'simple' performed better. At the intermediate scale there were no major differences between the two, and at the coarse scale GLM 'complex' was



Fig. 1. Observed and predicted distribution of *Quercus suber* in Catalonia. **A.** Observed distribution; **B.** Predicted distribution using classification tree analysis; **C.** Predicted distribution using generalized additive models.

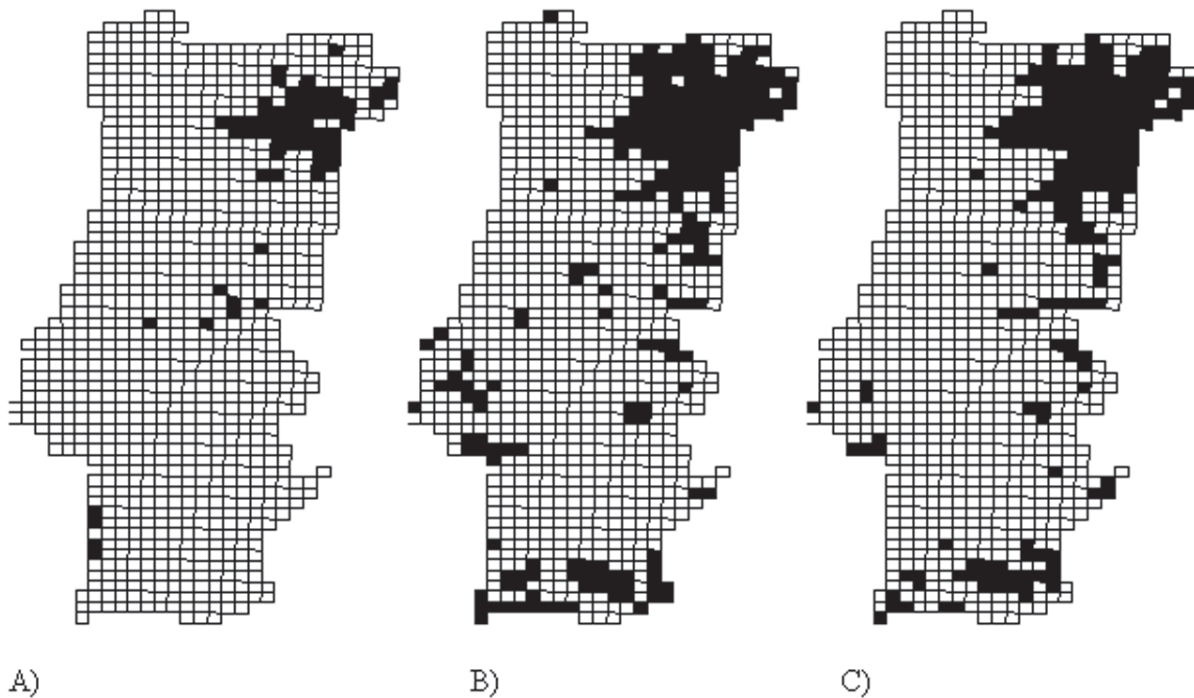


Fig. 2. Observed and predicted distribution of *Juniperus oxycedrus* in Portugal. **A.** Observed distribution; **B.** Predicted distribution using GLM 'complex'; **C.** Predicted distribution using GAM.

better. It appears that GLM requires increasing complexity as the scale of analysis increases, which could be interpreted in terms of species response shape to gradient. Being data-oriented and not influenced by a parametric choice, GAM is useful to explore response shapes (Oksanen & Minchin 2002). The response shape of *Quercus suber* to temperature of the coldest month in Catalonia showed a linear pattern (not presented). It is therefore logical that GLM 'simple', which uses only linear terms, provides as accurate a prediction as GLM 'complex' for this scale. Inversely, if we look at the coarsest scale, the response shape of *Quercus suber* to temperature of the warmest month was approximately a bell-shaped-curve (not presented). In this case, GLM with simple linear term cannot accurately fit the actual species response shape.

Generally, with our data, CTA performed poorly, but their discrimination ability increased with scale. At a finer scale, CTA had the lowest accuracy. At the intermediate scale CTA was nearly equal to the other methods, and at coarser scales CTA was similar to GLM 'simple'. The discrimination ability of GAM did not seem to be affected by changes in scale, suggesting that these kinds of models are relatively robust to scaling.

Species and environmental data

Although specific environmental variables differed across the three databases, they were still useful in

allowing simple comparisons to be drawn.

First, selected variables by models differed markedly between Catalonia/Europe and Portugal. Indeed, at fine and coarse scales, climatic variables were always selected as the main determinants of the distribution, whereas in Portugal, geologic and topographic variables were selected as being the most important. Although environmental data were highly related to topography in Portugal, if mean annual temperature and mean annual precipitation were the most important variables, models should have selected them as they did, for example, for *Pinus pinaster*.

In contrast, models for Catalonia and Europe were consistent in the variables they selected. Mean temperature of the coldest month was selected as the main factor for *P. pinaster*, *P. pinea* and *Q. suber* at both scales. In contrast, the variables selected by the models for *J. oxycedrus* were quite different for Catalonia, where precipitation was the most important variable, and Portugal or Europe where mean temperature of the warmest month was the key factor. For the European wide resolution, annual precipitation was never selected as one of the main factors determining species distributions, while mean of the coldest month and annual potential evapotranspiration were the most frequently selected variables.

Discussion

Effects of scale on model selection

The discriminative ability and predictive accuracy of our four modelling techniques showed different patterns as a function of the scale in the data. GAM provided fairly constant results across scales. A strong feature is that GAM is non-parametric and can take any smooth shape (Oksanen & Minchin 2002). This is particularly important as species response shapes are expected to change with scale, particularly for species with wide distributions (e.g. *Quercus suber*). In Catalonia, *Q. suber* has ranges that extend beyond the study area (Fig. 2) and thus the shape of response to a large climate gradient is likely to be monotonic as it only partially reflects the complete realized niche of the species. Hence, at this scale, GLM including only linear terms can be suitable to fit species response shapes. Conversely, at coarse scales, the shape of the response function describes a broader proportion of the realized niche. As shown here for *Q. suber*, and demonstrated by others, the shape of species response is likely to be Gaussian or asymmetric but not linear (Austin & Gaywood 1994; Austin & Meyers 1996; Huisman et al. 1993; Oksanen & Minchin 2002). GLM including more complex functions than linear are more suitable to fit species response shapes as the proportion of the realized niche included in the study increases.

Classification trees provide a different framework to model and predict species responses to environmental gradients. Our analyses showed that at a fine scale CTA had a lower accuracy than at intermediate and coarse scales when compared to generalized models. However, minimizing deviance and maximizing prediction accuracy may be conflicting goals. Indeed CTAs never performed better than generalized models at any of the scales, even if CTA yielded the lowest deviance (Franklin 1998). This lower power of CTA models may be attributable to the lack of interactive effects among environmental variables (Hastie & Tibshirani 1990). The increasing power of CTA with increasing scale may reflect the emergence of more complex relationships among variables. Nevertheless CTA remains an accurate method that makes it possible to describe how species respond to environmental variables, and to summarize species-environment relationships in a readable way.

We analysed real data to compare the models. Hence we cannot know the absolute performance of selected methods because 'truth' remains unknown. To test for absolute performance we could have utilized artificial data. However, results would be too much dependent on the assumptions used to generate the data and if they were unrealistic, then results could also be misleading.

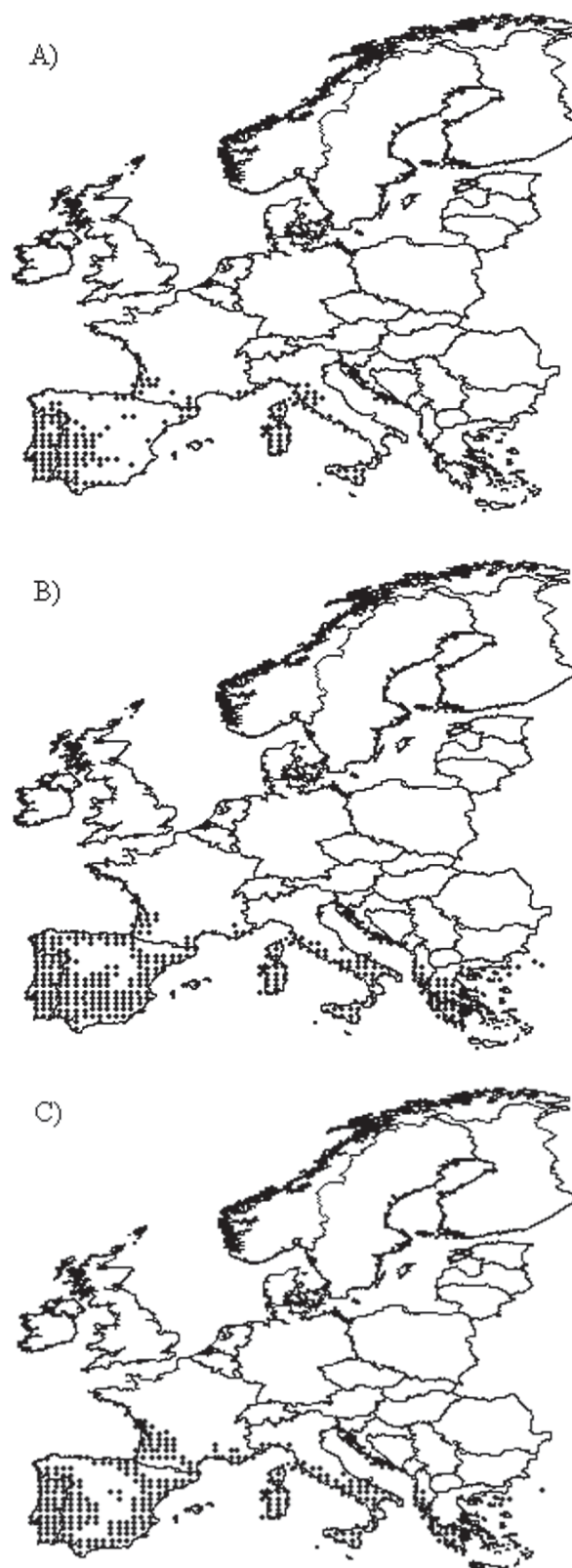


Fig. 3. Observed and predicted distribution of *Quercus suber* in Europe. **A.** Observed distribution; **B.** Predicted distribution using GLM 'simple'; **C.** Predicted distribution using GAM.

Factors influencing model fitting

There are many other parameters that influence differences in model adequacy across scales. First this study has shown that species with compact distributions are more easily modelled than species with scattered distributions. This phenomenon has already been documented (Araújo & Williams 2000) and relates to the previous point concerning the amplitude of gradients captured by models. Indeed, even at a fine scale, a species with a compact distribution is likely to have a greater proportion of its range within the sampled area. It is therefore appropriate to use such models to examine the shape of the response function describing the realized niche or predict range changes (Franklin 1998). Here, at fine and intermediate scales, species with compact distributions were well predicted by all models (*Q. suber* in Catalonia and *J. oxycedrus* in Portugal) and it is easy to explain why GLM 'simple' and GLM 'complex' had similar discrimination abilities.

Second, the assumption of correlative methods that species distributions are at equilibrium with their environment is a questionable one (Guisan & Zimmermann 2000). Non-equilibrium distributions resulting from history, biotic interactions, disturbances, random perturbations and human interferences can make predictions hazardous (Bolliger et al. 2000). Yet, impacts of forest management and other human activities, which do influence current plant distributions, remain hard to evaluate (Duckworth et al. 2000; Schulze & Kunz 1995). For instance, in Catalonia *Pinus pinea* and *P. pinaster* are intensively planted and their distributions are likely not to be at equilibrium with current environmental conditions, possibly explaining why models explained their distributions so poorly (Thuiller et al. in press). However, Bolliger et al. (2000) argues that at the regional scale correlative models are robust in predicting species even if they are not exactly at equilibrium. They also show that equilibrium models performed as well as dynamical models including physiological parameters and life history in predicting regional patterns of forest composition. This is still an unresolved debate.

Relevant environmental variables

Several authors have demonstrated the importance of climate on plant distributions at regional and continental scales (Woodward 1987, 1990), and that these are the result of specific physiological mechanisms (Woodward & Williams 1987). For example, annual or monthly minimum temperatures limit plant distributions by exceeding species lethal threshold for survival, while precipitation or water balance affect plant growth through control on leaf mass. In the Catalanian and European

databases, the main determinants of species distribution conformed to this model, and the mean temperature of the coldest month was a key explanatory variable for three of four species at both scales. In addition, seasonal precipitation was a better predictor of species distributions in Catalonia than mean annual precipitation (MAP), while in Europe mean potential evapotranspiration was a better predictor. Large intra-annual heterogeneity in rainfall in Europe and particularly in the Mediterranean is the likely reason why seasonal variables are more relevant to plant growth than MAP.

In contrast, the availability of climatic data for Portugal was limited to MAP and mean annual temperature (MAT), but contrary to the other data sets many other indirect variables such as topographic features were available. Models almost never selected MAT, whereas MAP was selected for all four species. Instead, several indirect variables such as slope and elevation were selected. The generally poor fits obtained suggest that plant distributions cannot be predicted accurately using a collection of indirect variables and only few simple climatic variables. Even when indirect variables are important locally, they may be applied only within a limited geographical extent without significant error, because in a different region the same topographic position may correspond to a different combination of direct and resource factors (Guisan & Zimmermann 2000). As a consequence, models based on indirect environmental variables have limited predictive power and cannot be used to project distributions elsewhere, or to model range shifts according to global change scenarios.

Alternatively, the poor accuracy of models at the intermediate scale (i.e. the Portuguese data set) may have resulted from the influence of human impacts on plant distributions. Using a database comprising 39 additional variables related to land use and land cover (<http://www.dga.pt>) we calculated new models (not presented) for the four species. Results were significantly better for *Pinus pinaster* (AUC evaluation = 0.85) and *Juniperus oxycedrus* (AUC evaluation = 0.96) only, suggesting that at this particular scale and for this data, species distributions are influenced by a combination of climatic, topographic and land use variables.

At the fine and intermediate scales, competitive interactions could also explain part of the spatial distribution of species. For example, Leathwick & Austin (2001) showed the importance of competition in explaining the spatial structure of forests in New Zealand. They argued that the abundance of *Nothofagus* species decrease the abundance of other conifer and broad-leaved tree species. They propose a framework, using GAM models and incorporating abundance of *Nothofagus* as independent variable to predict conifer and broad-leaved tree species. In our study, competitive

interactions were not directly taken into account. However, it is likely that the inclusion of information for other species as independent variables could improve the fit of the models. Arguably, however, it would be difficult to address competitive interactions of trees at the European scale using a resolution of 50 km × 50 km. Most plants occurring together in such grid cells will not compete with each other.

Extrapolating current and future species distributions

GLM 'complex' and GAM models were the most robust and accurate methods to interpolate current plant distributions at a coarse scale (Europe). Some authors have used non-parametric methods to predict future distributions from three bioclimatic variables (mean temperature of the coldest month, growing degree-days and ratio actual evapotranspiration/ potential evapotranspiration, including local weighted regression (loess) (Huntley et al. 1995; Beerling 1993) and cubic splines (Flannigan & Woodward 1994). Although robust non-parametric methods (GAM, loess) are data dependent and therefore lack underlying theoretical models, their applicability beyond the initial range for which they were developed remains questionable (but see Beerling et al. 1995). Alternatively, a framework using GAM to identify and test current response functions, and then deriving explicit functions to parameterize a GLM (Yee & Mitchell 1991), could be applied to extrapolate current species distributions to other areas or into the future under global change scenarios.

An important problem could arise in attempting to project future distributions, as the realized niche of species is dependent on abiotic factors, but also on biotic factors and particularly on competitive interactions. In the future, the realized niches may change as a result of competition from different species. However, if the modelling objective is to project future distribution of particular species or communities during the early stages of climate change, and if the site is dominated by competitors familiar to the target species, then the realized niche incorporating the average effects of interspecific competition across species' range may be used to a first approximation to future distribution.

Acknowledgements. This research is funded by the European Union (Advanced Terrestrial Ecosystem Analysis and Modelling project EVK2-CT-2000-00075. Species and environmental data for Portugal were kindly supplied by Pedro Segurado, University of Évora; The Centre for Ecological Research and Applied Forestry in Barcelona kindly supplied the species and climatic data for Catalonia. We thank Lluís Brotons, Mathieu Rouget, Guy Midgley, and two anonymous referees for their useful comments on the manuscript.

References

- Ahn, C.H. & Tateishi, R. 1994. Development of a global 30-minute grid potential evapotranspiration data set. *J. Jpn. Soc. Photogram. Rem. Sens.* 33: 12-21.
- Araújo, M.B. & Williams, P. 2000. Selecting areas for species persistence using occurrence data. *Biol. Conserv.* 96: 331-345.
- Austin, M.P. 1985. Continuum concept, ordination methods, and niche theory. *Annu. Rev. Ecol. Syst.* 16: 39-61.
- Austin, M.P. & Gaywood, M.J. 1994. Current problems of environmental gradients and species response curves in relation to continuum theory. *J. Veg. Sci.* 5: 473-482.
- Austin, M.P. & Meyers, J.A. 1996. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *For. Ecol. Manage.* 85: 95-106.
- Austin, M.P. & Smith, T.M. 1989. A new model for the continuum concept. *Vegetatio* 83: 35-47.
- Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. 1994. Determining species response functions to an environmental gradient by means of a beta-function. *J. Veg. Sci.* 5: 215-228.
- Bakkenes, M., Alkemade, R.M., Ihle, F., Leemans, R. & Latour, J.B. 2002. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biol.* 8: 390-407.
- Barbéro, M., Loisel, R., Quézel, P., Richardson, D.M. & Romane, F. 1998. Pines of the Mediterranean Basin. In: Richardson, D.M. (ed.) *Ecology and biogeography of Pinus*, pp. 450-473. Cambridge University Press, Cambridge, UK.
- Beerling, D. J. 1993. The impact of temperature on the northern distribution limits of the introduced species *Fallopia japonica* and *Impatiens glandulifera* in north-west Europe. *J. Biogeogr.* 20: 45-53.
- Beerling, D.J., Huntley, B. & Bailey, J.P. 1995. Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surface. *J. Veg. Sci.* 6: 269-282.
- Bio, A.M.F., Alkemade, R. & Barendregt, A. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. *J. Veg. Sci.* 9: 5-16.
- Bolliger, J., Kienast, F. & Bugmann, H. 2000. Comparing models for distributions: concept, structures, and behavior. *Ecol. Model.* 134: 89-102.
- Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. 1984. *Classification and regression trees*. Chapman and Hall, New York, NY, US.
- Burke, A. 2001. Classification and ordination of plant communities of the Naukluft Mountains, Namibia. *J. Veg. Sci.* 12: 53-60.
- Carey, P.D. 1996. A cellular automaton for predicting the distribution of species in a changed climate. *Global Ecol. Biogeogr. Lett.* 5: 217-226.
- Chambers, J.M. & Hastie, T.J. 1997. *Statistical models in S*. Chapman & Hall, London, UK.
- Duckworth, J.C., Bunce, R.G.H. & Malloch, A.J.C. 2000. Modelling the potential effects of climate change on calcareous grasslands in Atlantic Europe. *J. Biogeogr.* 27: 347-358.
- Eastman, J.R. 1996. *Idrisi for Windows – User's Guide Ver-*

- sion 2.0. (Clark Labs for Cartographic Technology and Geographic Analysis) Clark University, Worcester, MA, US.
- Fielding, A.H., & Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24: 38-49.
- Flannigan, M.D. & Woodward, F.I. 1994. Red pine abundance: current climatic control and response to future warming. *Can. J. For. Res.* 24: 1166-1175.
- Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *J. Veg. Sci.* 9: 733-748.
- Guisan, A. & Theurillat, J.-P. 2000. Equilibrium modelling of alpine plant distribution: how far can we go? *Phytocoenologia* 30: 353-384.
- Guisan, A. & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135: 147-186.
- Guisan, A., Theurillat, J.-P. & Kienast, F. 1998. Predicting the potential distribution of plant species in an alpine environment. *J. Veg. Sci.* 9: 65-74.
- Guisan, A., Weiss, S.B., & Weiss, A.D. 1999. GLM versus CCA spatial modelling of plant species distribution. *Plant Ecol.* 143: 107-122.
- Hanley, J.A. & McNeil, B.J. 1983. A method of comparing the areas under receiver operating characteristic (ROC) curve. *Radiology* 148: 839-843.
- Hastie, T.J. & Tibshirani, R. 1990. *Generalized additive models*. Chapman and Hall, London, UK.
- Huisman, J., Olff, H. & Fresco, L.F.M. 1993. A hierarchical set of models for species response analysis. *J. Veg. Sci.* 4: 37-46.
- Huntley, B. 1995. Plant species' response to climate change: implications for the conservation of European birds. *IBIS* 137: 127-138.
- Huntley, B., Berry, P.M., Cramer, W. & McDonald, A.P. 1995. Modelling present and potential future ranges of some European higher plants using climate response. *J. Biogeogr.* 22: 967-1001.
- Huston, M.A. 1994. *Biological diversity*. Cambridge University Press, Cambridge, UK.
- Iverson, L.R. & Prasad, A. 1998. Predicting abundance for 80 tree species following climate change in the Eastern United States. *Ecological Monographs* 68(4): 465-485.
- Jalas, J. & Suominen, J. 1972-1996. *Atlas Florae Europaeae*. Vol. 1-11. Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki, FI.
- Lahti, T. & Lampinen, R. 1999. From dot maps to bitmaps – Atlas Florae Europaeae goes digital. *Acta Bot. Fenn.* 162: 5-9.
- Leathwick, J.R. & Austin, M.P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* 82: 2560-2573.
- Legates, D.R. & Wilmott, C.J. 1990a. Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.* 10: 111-127.
- Legates, D.R. & Wilmott, C.J. 1990b. Mean seasonal and spatial variability in global surface air temperature. *Theor. Appl. Climatol.* 41: 11-21.
- McCullagh, P. & Nelder, J.A. 1989. *Generalized linear models*. Chapman & Hall, London, UK.
- Midgley, G.F., Hannah, L., Millar, D., Thuiller, W. & Booth, A. 2003. Developing regional and species-level assessments of climate change impacts on biodiversity: A preliminary study in the Cape Floristic Region. *Biol. Conserv.* 112: 87-97.
- Nicholls, A.O. 1989. How to make biological surveys go further with generalized linear models. *Biol. Conserv.* 50: 51-75.
- Ninyerola, M., Pons, X. & Roure, J.M. 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *Int. J. Climatol.* 20: 1823-1841.
- Oksanen, J. & Minchin, P.R. 2002. Continuum theory revisited: what shape are species responses along ecological gradients. *Ecol. Model.* 157: 119-129.
- Olden, J.D. & Jackson, D.A. 2002. Illuminating the 'black box': a randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Model.* 154: 16.
- Pearce, J. & Ferrier, S. 2000a. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* 133: 225-245.
- Pearce, J. & Ferrier, S. 2000b. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecol. Model.* 128: 127-147.
- Pearson, R.G., Dawson, T.P., Berry, P.M. & Harrison, P.A. 2002. SPECIES: A spatial evaluation of climate impact on the envelope of species. *Ecol. Model.* 154: 289-300.
- Schulze, R.E. & Kunz, R.P. 1995. Potential shifts in optimum growth areas of selected commercial tree species and subtropical crops in southern Africa due to global warming. *J. Biogeogr.* 22: 679-688.
- Swets, K.A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285-1293.
- ter Braak, C.J.F. 1987. The analysis of vegetation-environment relationship by canonical correspondence analysis. *Vegetatio* 69: 69-77.
- Thuiller, W., Vaydera, J., Pino, J., Sabaté, S., Lavorel, S. & Gracia, C. 2003. Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecol. Biogeogr.* 12: 313-325.
- Vayssières, M.P., Plant, R.E. & Allen-Diaz, B.H. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *J. Veg. Sci.* 11: 679-694.
- Vetaas, O.R. 2000. Comparing species temperature response curves: population density versus second-hand data. *J. Veg. Sci.* 11: 659-666.
- Woodward, F.I. 1987. *Climate and plant distribution*. Cambridge University Press, Cambridge, UK.
- Woodward, F.I. 1990. The impact of low temperatures in controlling the geographical distribution of plants. *Phil. Trans. R. Soc. Lond. B* 326: 585-593.
- Woodward, F.I. & Williams, B.G. 1987. Climate and plant distribution at global and local scales. *Vegetatio* 69: 189-197.
- Yee, T.W. & Mitchell, N.D. 1991. Generalized additive models in plant ecology. *J. Veg. Sci.* 2: 587-602.

Received 11 April 2002;

Revision received 15 January 2003;

Accepted 13 February 2003.

Coordinating Editor: N. Kenkel.

ANNEXE 1.2

Evaluating species distribution models: methodologies and conservation implications of prevalence and probability thresholds

Thuiller W., Howell C.A., Midgley G.F., Rouget M.

Soumis

Evaluating species distribution models: methodologies and conservation implications of prevalence and probability thresholds

Wilfried Thuiller¹, Christine A. Howell², Guy F. Midgley^{3,4} and Mathieu Rouget⁵

¹Centre d'Ecologie Fonctionnelle et Evolutive. Centre National de Recherche Scientifiques. 1919 route de Mende, 34293 Montpellier cedex 5, France

²Department of Biology & International Center for Tropical Ecology, University of Missouri-St. Louis, 8001 Natural Bridge Rd., St. Louis, MO 63121-4499, USA

³Climate Change Research Group, Kirstenbosch Research Center, National Botanical Institute, P/Bag x7, Claremont 7735, Cape Town, South Africa.

⁴Center for Applied Biodiversity Science, Conservation International, 1919 M St., NW, Washington, D.C. 20036

⁵Institute for Plant Conservation, Department of Botany, University of Cape Town, Private Bag Rondebosch 7701, South Africa

Corresponding author: Wilfried Thuiller, E-mail: wilfried.thuiller@cefe.cnrs-mop.fr

Fax: 33 (0)4-67-41-21-38

Number of words: 6439

Running title: *EVALUATING SPECIES DISTRIBUTION MODELS*

ABSTRACT

Climate and land-use are changing rapidly, with potentially significant consequences for the regional distribution and persistence of species. Because biodiversity is closely linked to habitat and climate at the regional scale, there is an increasing interest in modelling the regional distributions of individual species, and many methods are now available. However, relatively little attention has been given to assessing reliability of model predictions. Here we analyse three methods (ROC curve, Kappa statistic and explained deviance) to evaluate the performance of logistic regression applied to occurrence data for species with varying prevalence (common to rare). Although deviance is often used to assess the accuracy

of species distribution models, we find it to be inadequate. We also find that ROC curve and Kappa have different behaviour according to species prevalence (rarity) and are often in disagreement. We discuss the implications of ROC curve and Kappa statistic for applied questions in ecology and conservation biology and conclude that choice of evaluation method should be based on careful examination of goals and data, rather than on subjective preference of the user.

KEYWORDS: Kappa statistic, Modelling occurrences, Predictive performance, ROC curve, Species conservation.

INTRODUCTION

With the development of GIS and spatial statistical software, the prediction of species' distributions from environmental data using correlative methods is increasingly used (Araújo, 1999; Peterson & Vieglais, 2001, Thuiller et al., In review-b). Relationships linking species with their abiotic environment are examined in order to model the current distribution of species (Chamberlain et al., 1999; Cowley et al., 2000; Green et al., 1994), to forecast the potential distribution of species (Baker et al., 2000; Guisan et al., 1998; Sykes et al., 1996) and to project the species' future distributions under projected global climate change (Beerling, 1993; Sykes et al., 1999) or invasion into a novel environment (Rouget et al., 2001). Such modelling exercises often apply general linear models

or related techniques to examine the presence or absence of a species from a set of survey sites or atlas data in relation to environmental variables, thereby making it possible to predict the probability of occurrence of the species at unsurveyed or unsampled sites.

Many statistical approaches are available to predict species' distributions with presence-absence data, including generalized linear models (McCullagh and Nelder, 1989), generalized additive models (Hastie and Tibshirani, 1990), or classification and regression tree analysis (Breiman et al., 1984). These models produce a probability surface based on the correlation between species distributional data and abiotic factors. Once the resulting probability surface is modelled there are two additional considerations in the modelling process:

model evaluation and the selection of a probability threshold.

Evaluating the predictive performance of the resulting models is a critical step in model development and acceptance. Model evaluation determines the suitability of using a model for a specific purpose (Buckland and Elston, 1993) alone or compared to alternative modelling techniques. Despite the growing use and acceptance of predictive models, there have been few studies examining the methods used to evaluate the predictive performance of models (but see Fielding and Bell, 1997; Manel et al., 2001; Pearce and Ferrier, 2000). Ideally, evaluation should be based on independent data (i.e. data not used to construct the model). If independent data are unavailable, as is often the case, then statistical resampling including jack-knifing or cross-validation can be used (Manel et al., 2000). Alternatively the initial data can be randomly subdivided into calibration and evaluation datasets. The calibration data is used to run the model and the evaluation data is used to assess model performance. With any of these evaluation approaches the species' prevalence (ratio of total number of presences against total number of areas considered in the models) may influence evaluation methods and the nature of the resulting prediction errors (but see Manel et al., 2001). A final consideration in the modelling process is the transformation of the modelled probability values into binary presence/absence values based on a chosen probability threshold. For some projects this will not be a necessary step as the original probability value is retained in the final product (e.g. Araújo et al., In press), but in other cases a binary presence/absence map is required. Significant problems may arise in deciding on the probability thresholds that are used to map species presence or absence. This can be seen as a balance between "liberal" and "conservative" approaches that have rather different implications in practice. A liberal approach sets probability threshold at low levels, therefore maximizing the chance of identifying locations where the species may be located, but potentially including areas in which the target taxa is not located. However, a more conservative, high cut-off value will select locations with a high probability of containing the target taxa, but may also simultaneously exclude other locations where the target taxa occurs.

This paper addresses the two aforementioned modelling steps following model construction: model evaluation and the selection of a probability threshold. As we demonstrate, these two steps may be interrelated. We analyse the response of three commonly used evaluation methods (ROC curve, Kappa statistic, and adjusted-D²) to the prevalence of the species modelled and the nature of prediction errors generated; we then explore the statistical and conservation implications of different probability thresholds. Our specific objectives are to: 1) Briefly describe threshold-dependent and threshold-independent model evaluation methods; 2) Present a

case study in which we examine the relationship between prevalence and model evaluation statistics for 50 South African plant species; 3) Quantify how threshold determination affects our results from the previous objective; 4) Consider the relevance and importance of prevalence and various prediction errors in the application of predictive models to conservation planning, invasion and reintroduction studies, and forecasting future distributions under global climate change.

MATERIALS AND METHODS

Methods for evaluating model performance: a short review

Few methods are available to evaluate the performance of spatial predictive models (see Fielding and Bell, 1997 for a review). Two methods are generally used in ecological studies: the relative operating characteristic curve (ROC curve) (Cumming, 2000; Pearce and Ferrier, 2000) and the Kappa statistic (Collingham et al., 2000; Monserud and Leemans, 1992). These methods are based on the cross-tabulation of the observed and predicted patterns, summarized in a misclassification matrix (Table 1). Two types of errors are measured: false negative (observed presences not correctly predicted: FN) and false positive (observed absences not correctly predicted: FP). Therefore two indices in percentages are commonly defined from the true positives (observed presences correctly predicted: TP) and the true negatives (observed absences correctly predicted: TN): sensitivity (or true positive fraction: $se = TP/(TP+FN)$) and specificity (true negative fraction: $sp = TN/(TN+FP)$). They measure the proportion of sites or cells at which the observations and the predictions are in agreement. Moreover the error, sensitivity and specificity can be related: FP fraction = 1- specificity and FN fraction = 1- sensitivity.

Fielding & Bell (1997) and Manel et al. (2001) described three values derived directly from the misclassification matrix and used in model evaluation: the odds ratio, the normalized mutual information statistic (NMI) and the Kappa statistic. These measures provide an assessment of the rate of correctly predicted presences expected by chance. However, the first two are rarely used because of discrepancies related to their calculation mode (for more details see Fielding and Bell, 1997; Manel et al., 2001). The Kappa statistic (Cohen, 1960), which is the most commonly used method in ecology (Manel et al., 2001), consists in the proportion of specific agreement (see Monserud and Leemans, 1992 for the algorithm)(Table 2). However, Fielding & Bell (1997) and Forbes (1995) note that Kappa is sensitive to sample size and fails when the size of one class greatly exceeds any other. Thus if the species has a low prevalence (which often occurs with rare species) or a high prevalence in the sample, the Kappa statistic can be misleading. However, in a recent study Manel et al. (2001) showed that Kappa is not affected by the prevalence. We address this contradiction in this paper.

All these measures including the Kappa statistic depend on a threshold criterion applied *a priori* to the probability values generated by the models. Commonly this threshold aims to convert the probability values into binary form (0 for absence and 1 for presence). *A priori* choice of threshold is problematic because it directly influences the misclassification matrix, the evaluation indices, and associated interpretations; yet this issue is seldom addressed in model evaluation studies.

Different ways to select an “optimum probability threshold” are used: arbitrary threshold (Franklin, 1998; Midgley et al., In press), crossover of the misclassification rates (sensitivity cross specificity) (Cumming, 2000; Pearce et al., 2001) or the threshold maximizing the Kappa statistic (Collingham et al., 2000; Guisan et al., 1998).

Another method to evaluate predictive performance of models is the receiver operating characteristic curve (ROC curve). This approach has frequently been used in medicinal studies (Forbes, 1995; Hanley and McNeil, 1982, 1983) although there are recent applications in ecological-related studies (Cumming, 2000; Pearce and Ferrier, 2000; Thuiller et al., In review-a). It is based on the graphical representation of the sensitivity as a function of 1 minus specificity (true positive fraction versus false positive fraction) for varying thresholds. This approach provides a straightforward way to estimate the predictive performance of the model because it uses the probability values of different models (e.g. logistic regression, generalized additive models, decision tree, artificial neural networks) without converting to binary form based on a subjective threshold. With the observed values and the predicted values, it is simple to compute the ROC curve and so evaluate the predictive performance. Hence, if the representation of binary distribution or the computation of the misclassification matrix is required the same methods as Kappa can be used to determine the threshold; alternatively one could represent the results on a map using the actual probabilities of occurrence.

Good accuracies are characterized by high values of sensitivity and low values of 1-specificity, and can be measured by the area under the curve (AUC) that is above a 45° line where the axes of the ROC curve are sensitivity versus 1-specificity calculated for a range of thresholds (Fig. 1a & Table 2). Similarly we can plot sensitivity and specificity versus a range of thresholds (Fig. 1b). In fact, the threshold corresponding to the crossover of sensitivity and specificity corresponds to the inflection point of the ROC curve (Fig. 1c). These three figure panels have the same information but use three different representations. Manel et al. (2001) argue that the ROC evaluation method is influenced by prevalence, and tends to over-estimate the predictive performance of models with low prevalence.

The proportion of explained deviance can also be used as evaluation method (Guisan et al., 1999), that is a similar measure as the variance (McCullagh and Nelder, 1989). This measure can be corrected by the number of degree of freedom used to fit the logistic regression, providing an adjusted-D² (Guisan et al., 1999). This provides a measure of goodness of fit of the model and allow quick comparison between nested models.

Case study: data source, study area and methods

To analyse different evaluation methods, we chose to use biological rather than theoretical data. We examined the three previous methods to evaluate predictive performance of models: first the reliability of the goodness of fit derived from a generalized linear or additive models (adjusted-D²), second the ROC curve method applied to derive AUC, and then to identify an optimal threshold, and finally the Kappa statistic. We used these methods to evaluate the performance of a generalized linear model (the most commonly used method to predict species distribution) on 50 Proteaceae species from South Africa. Results from this study could be extrapolated to other types of models used in biogeography studies, like generalized additive models (Thuiller et al., In review-a; Yee and Mitchell, 1991), artificial neural networks (Manel et al., 1999; Spitz and Lek, 1999), decision-trees (Rouget et al., 2002; Thuiller et al., In review-b).

These 50 species are present in the Protea Atlas Database (Rebello, 1995), which contains records of both species presence and absence at more than twenty-thousand points within the Fynbos Biome in the Cape Floristic Region. Seven climatically derived parameters considered critical to plant physiological function and survival (potential evapotranspiration, mean annual temperature, minimum T° of the coldest month, annual, winter and summer precipitation and annual sum of daily T° greater than 18°C) and three soils variables were used (index of soil fertility, percentage of clay and sand) (Schulze, 1997).

Seven prevalence values ($p = 0.05; 0.1; 0.25; 0.5; 0.75; 0.9; 0.95$) for each species were derived by randomly removing absences from the total dataset to reach the desired prevalence (the initial number of presences for each species was always conserved). The fifty selected species have approximately the same number of occurrences in the Atlas (200-500 occurrences on total initial cells). We modelled each species using logistic regression for each prevalence value and calculated the corresponding adjusted-D², the ROC curve and AUC, as well as the Kappa statistic (and misclassification matrix) for varying thresholds. Threshold determination followed four methods: first using an arbitrary threshold (0.5), second using the prevalence as threshold (depending on the species and the prevalence), third using the “optimal” threshold maximizing the Kappa, and fourth using the cross-over of sensitivity-specificity (ROC approach as depicted in Fig. 1c).

RESULTS

Threshold independent predictive performance:

We first consider the overall predictive performance for evaluation methods that do not require data transformation from probability values into binary presence-absence values.

Fig. 2 shows the area under the curve and the adjusted- D^2 as a function of the prevalence of the species. The curves have distinctive patterns. The AUC is rather stable (Fig. 2a) with only a slight increase with prevalence, whereas goodness-of-fit exhibits a decrease (Fig. 2b). The lack of variability of the AUC with different prevalence values shows the stability of this approach for the 50 different species. The ROC curve appears to be a stable method when prevalence varies. The low variability shows that results are not biased by the species selected (Fig. 2a). The method responds equally well with all species.

Threshold-dependent predictive performance:

We next consider threshold-dependent predictive performance for those approaches where one must first determine a specific threshold value in the misclassification matrix. We analysed the response of the Kappa statistic to a range of prevalence and threshold values (Fig. 3). The general pattern of the Kappa statistic is very similar among the different methods for selecting the threshold. However, two groups of similar patterns appear: first, Kappa calculated with an arbitrary threshold concurs with the one calculated using a threshold derived from a maximisation of Kappa (Fig. 3a & 3c), and second, Kappa calculated with a threshold equal to the prevalence of the species concurs with the one using threshold determined by the crossover of sensitivity and specificity (ROC approach) (Fig. 3b & 3d). The second group is distinguished by increased variability at high prevalence values compared to the low variability within the first group.

Naturally, the best evaluation of models is provided by the Kappa with threshold maximizing the Kappa. Another point to notice is that the previous Kappa (Fig. 3c) shows similar pattern of the AUC (Fig. 2a) with a stronger increase following the threshold. Like the AUC, the variability of the results is greater for low prevalence and very low for high prevalences. This result is intuitive because with a high prevalence the correlative methods show good predictions whatever the species and bioclimatic variables are. However, for low prevalence values, species' response to bioclimatic variables cannot clearly be discerned due to a low sample size of occurrences, and thus, for low prevalence the variability of predictive performance is greater. Even if AUC and Kappa show similar patterns (Fig. 2a & 3c), results of predictive performance differ somewhat, especially at low prevalence (Table 2). When prevalence equalled 0.05 or 0.1, the AUC exhibited very good predictive performance ($AUC > 0.9$) whereas the Kappa statistic (Fig. 3c) exhibited relatively good-fair predictive performance ($0.4 < K < 0.6$). It is only when prevalence

equals to 0.5 or 0.75 that both evaluators show similar results.

Impact of threshold and prevalence: To further understand why there are differences between both methods and to better explain the impact of the threshold on the misclassification matrix, we analysed the sensitivity and specificity following the method used to optimise of threshold for each species and each prevalence. The impact of the threshold on the misclassification matrix is shown in Fig. 4. Paired-wise graphics indicate different patterns following the method used to choose the threshold for sensitivity and specificity. Threshold determination methods include maximization from Kappa, maximization from ROC, arbitrary threshold equal to 0.5, and threshold equal to the prevalence in the data. As expected, the threshold had an impact on sensitivity (*se*), specificity (*sp*) and therefore on the associated errors. Again, graphics for threshold equals prevalence (Fig. 4g & 4h) and threshold from the ROC approach (Fig. 4c & 4d) show similar values. Both *se* and *sp* are always high and approximately equal, even if the threshold by ROC approach shows less extreme values. The graphics with threshold by maximization of Kappa (Fig. 4a, b) and threshold=0.5 (Fig. 4e & 4f) exhibit a similar pattern where the sensitivity is low for low prevalence and the specificity is low for high prevalence. However, marked differences are visible for low prevalence where the sensitivity is very low with the arbitrary threshold (0.5).

If we compare the sensitivity and specificity for the most commonly used methods (threshold by ROC approach and threshold by maximization of the Kappa) as a function of prevalence, further aspects of model evaluation are uncovered. We consider an example where we investigate the misclassification matrices for both the ROC and Kappa methods and using two extreme values for prevalence (Table 3 and Table 4).

For the high prevalence (prevalence = 0.95) and threshold derived from ROC approach, the misclassification matrix exhibits both sensitivity and specificity around 92%, whereas with threshold determined by maximization of the Kappa, *se* equals 79.1% and *sp* equals 96.9% (Table 3). However, with the counts, we can see that the number of errors with the first method is greater (33 errors) than the second (4 errors). The first method gives a high number of false negatives. The model predicts more than twice the absences than in the reality.

For low prevalence (threshold = 0.05), we found the same pattern (Table 4). However, but choosing the best method remains more complicated. With the ROC approach, there are many errors (549 errors) compared with the maximization by Kappa (350 errors). Typically the best method depends on the nature of the study, because the method that maximizes Kappa gives relatively more false negatives (97) than the ROC curve approach (27; Table 4). However, overall the model results based on maximizing Kappa are often

very similar to the AUC approach, showing that these indices are not that different.

DISCUSSION

It is important to comprehend the relationship between predictive accuracy and prevalence for a variety of ecological as well as statistical reasons. Statistically, low prevalence values indicate that absence data far exceeds presence data. Ecologically, low prevalence values are generally associated with rare species. These are often the species conservationists are most interested in modelling and are frequently those of greatest environmental concern (Midgley et al., In press; Peterson et al., 2001). Similarly, high prevalence values indicate statistically that presence data exceeds absence data. Ecologically high prevalence values are often, but not always associated with common or more abundant species. A prevalence value cannot necessarily be used to determine whether a species is rare or abundant, but specific prevalence values are often a property of rare and abundant species (Stockwell and Peterson, 2002). There are other reasons that a given dataset may contain a specific ratio of presences and absences including the vagaries of data collection. Often we know more about where rare species are not found, than where they are found. On the other hand, there are certainly museum datasets in which rare species are better represented than common species because rare species are highly sought by collectors. Because of this inconsistent association between prevalence and rarity it is important to consider the potential effects of prevalence on the modelling and model evaluation processes (Manel et al., 2001).

Model Evaluation: This study showed that the AUC derived from the ROC curve exhibits stability across prevalence values (Fig. 2), as suggested by Manel et al. (2001) and can be considered as a good overall evaluation method. Threshold-independent evaluators such as AUC are advantageous because one does not need to convert prediction probability values into binary form (Cumming, 2000). However, evaluation should be done within the ROC curve framework and not within the general goodness-of-fit (adjusted-D²) framework provided by some correlative models (e.g. null deviance method included in some statistical modelling software packages). Goodness-of-fit measures do not make an assessment of predictive performance (although their failure underscores aspects of the relationship between the modelled species distributions and the explanatory variables used in the model). Indeed, we saw that the adjusted-D² was very low when prevalence equals 0.95 whereas the percentage of sites well predicted is close to 95%. In fact, logistic regression may produce scores that are biased toward the larger group (Hosmer and Lemeshow, 1989). So, for a high prevalence, the probability values are close to one and thus close to the null model. This is misleading in suggesting that the addition of bioclimatic variables is not necessary and

that a null model is sufficient to predict the species distribution at a given prevalence. This particularity is highly problematic and highlights the fact that minimizing deviance is not equivalent to maximizing predictions. Clearly, this method should not be used for model evaluation.

Probability values from models may later be converted into binary form (if binary representation is required) using different threshold determination techniques (discussed below). However, there are many instances where the conversion of probability values is not required and results can be given in terms of actual probabilities for each locality of interest. Maintaining and representing model results in probability form also allows readers to form their own conclusions about the data, and it can be the most appropriate output form for certain conservation (Araújo and Williams, 2000) and invasion issues (Higgins et al., 1999). For example, a predictive model can be run to estimate the probability of species occurrence in a selected area in order to estimate the persistence of the species (Araújo et al., In press).

Another advantage of ROC evaluator is the possibility to compare several curves from different models (Hanley and McNeil, 1982, 1983) or from different studies. While Kappa is a reasonable evaluation method, it is not as flexible as ROC in comparative studies. Different algorithms such as generalized linear models, generalized additive models and classification trees may be reliably compared based on AUC (Thuiller *et al* in review). This comparison examines the difference between the AUC among models and provides a better method than general accuracy (percentage of cells correctly classified), which is highly biased by the sample size and the prevalence of the species (Fielding and Bell, 1997; Pearce and Ferrier, 2000). The AUC is a good method to compare results from different studies, because it is not dependent on prevalence, sample size, or most importantly, on the threshold used. For example, Bakkenes et al. (2002) compared their species distributions with Huntley et al., 1995's results using the Kappa values of both studies. Yet, no statistical test was performed, and no confidence intervals were used. Therefore, it is not appropriate or meaningful to compare Kappa results when the thresholds are likely different. The ROC curve approach avoids these drawbacks.

Threshold selection and prevalence: Recently Manel et al. (2001) highlighted the importance of prevalence on the determination of an appropriate threshold. Their general conclusion was that using the ROC approach to find the "optimum" threshold over-estimated true occurrence amongst scarcer species. They found that this method led to a false evaluation and that the "threshold by maximizing the Kappa" approach was safer. In fact, the problem is not only statistical but also ecological because conclusions based on model' results depend on conservation implications and other issues like biodiversity management of false negative

and false positive errors (the results of liberal and conservative assumptions) which should be dictated by the goal of the study.

For conservation studies in which the goal is to delimit the range of a threatened species to be protected, a liberal modelling threshold is necessary to minimize the percentage of false negative errors. This liberal approach will allow for the maximal amount of the potential habitat for the endangered species (Loiselle et al. in review). Threshold selection based on ROC (which minimizes the difference between sensitivity and specificity) is therefore a good method. Conversely for reintroduction studies, we need to ensure that the species can develop in the specified habitat. Only the most optimal habitat should be used (so as to not waste resources) and omission errors must to be minimized; therefore a very conservative model interpretation is justified (Lomolino and Channell, 1998). In this case thresholds derived from maximizing Kappa are preferable. For global climate change studies, the interpretation of results may vary considerably with liberal versus conservative modelling thresholds. It may therefore be warranted to provide both liberal and conservative scenarios when modelling global climate change and to clearly state the threshold determination method used.

The arbitrary threshold of 0.5 or threshold equals to prevalence are generally too extreme to provide good model assessment for any conservation scenario. However, we saw that for prevalence closed to 0.5, both thresholds could be used to avoid the calculation of crossover *se-sp* and maximisation of Kappa.

Additionally, we must consider that an absence in occurrence data is not always a real absence. Green (1971) provides three possible interpretations when a species is predicted as absent: "(1) the species cannot live there, that is, its niche does not include that point, (2) the species can live there, but never had the opportunity for zoogeographical reasons, and (3) the species can and does live there, but the sample failed, by chance, to include a representative of that species." So for species with low prevalence (scarce), over-estimating presences is not nonsense because we must consider that the predictions reveal the potential niche of the species and that it is not present now in these plots for many reasons, including low dispersion, land use factors and so on. Obviously this over-estimation must be controlled, and we cannot say that all areas predicted as present by models are potential habitat for the species because real errors may also occur. This problem must to be taken into consideration when determining the goals of the study.

To conclude, we find that prevalence could have effects on predictive models as well as on evaluation methods. However, to provide more ecological sense to evaluation, further studies about models' performance should accord importance to the kind of species distribution. Indeed, widespread and aggregated distributions, as well as species niche positions and tolerances could influence species

modelling and evaluation methods. Further developments could have to include this kind of information in the predictive models and evaluation methods. It could offer ecological sense to statistical outputs.

ACKNOWLEDGMENTS

This research was funded by the Ministère des Affaires Etrangères et Enseignement Supérieur of France and the National Research Foundation of South Africa. The first author is funded by the European Union's ATEAM (EVK2-CT-2000-00075) project. We are indebted to Sandra Lavorel, Lluís Brotons and Miguel Araújo for their comments on successive versions of the manuscript.

CITED REFERENCES

- Araújo, M. B., 1999. Distribution patterns of biodiversity and the design of a representative reserve network in Portugal. *Diversity and Distributions* 5, 151-163.
- Araújo, M. B., and Williams, P., 2000. Selecting areas for species persistence using occurrence data. *Biological Conservation* 96, 331-345.
- Araújo, M. B., Williams, S. E., and Fuller, R. J., In press. Dynamics of extinction and the selection of nature reserves. *Proceedings of the Royal Society London B*.
- Baker, R. H. A., Sansford, C. E., Jarvis, C. H., Cannon, R. J. C., MacLeod, A., and Walters, K. F. A., 2000. The role of climatic mapping in predicting the potential geographical distribution of non-indigenous pests under current and future climates. *Agriculture, Ecosystems and Environment* 82, 57-71.
- Bakkenes, M., Alkemade, R. M., Ihle, F., Leemans, R., and Latour, J. B., 2002. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology* 8, 390-407.
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., 1984. *Classification and regression trees*. Chapman and Hall, New York.
- Buckland, S. T., and Elston, D. A., 1993. Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology* 30, 478-495.
- Chamberlain, D. E., Wilson, A. M., Browne, S. J., and Vickery, J. A., 1999. Effects of habitat type and management on the abundance of skylarks in the breeding season. *Journal of Applied Ecology* 36, 856-870.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37-46.
- Collingham, Y. C., Wadsworth, R. A., Huntley, B., and Hulme, P. E., 2000. Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology* 37, 13-27.
- Cowley, M. J. R., Wilson, R. J., León-Cortés, J. L., Gutiérrez, D., Bulman, C. R., and Thomas, C. D., 2000. Habitat-based statistical models for

- predicting the spatial distribution of butterflies and day-flying moths in a fragmented landscape. *Journal of Applied Ecology* 37, 60-72.
- Cumming, G. S., 2000. Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography* 27, 441-455.
- Fielding, A. H., and Bell, J. F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24(1), 38-49.
- Forbes, A. D., 1995. Classification algorithm evaluation: five performance measured based on confusion matrix. *Journal of Clinical Monitoring* 11, 189-206.
- 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.
- Green, R. E., Osborne, P. E., and Sears, E. J., 1994. The distribution of passerine birds in hedgerows during the breeding season in relation to characteristics of hedgerows and adjacent farmlands. *Journal of Applied Ecology* 31, 677-692.
- Green, R. H., 1971. A multivariate statistical analysis approach to the Hutchinson niche: bivalve molluscs of central Canada. *Ecology* 52(4), 543-556.
- Guisan, A., Theurillat, J.-P., and Kienast, F., 1998. Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science* 9, 65-74.
- Guisan, A., Weiss, S. B., and Weiss, A. D., 1999. GLM versus CCA spatial modelling of plant species distribution. *Plant Ecology* 143, 107-122.
- Hanley, J. A., and McNeil, B. J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29-36.
- Hanley, J. A., and McNeil, B. J., 1983. A method of comparing the areas under receiver operating characteristic (ROC) curve. *Radiology* 148, 839-843.
- Hastie, T. J., and Tibshirani, R., 1990. *Generalized additive models*. Chapman and Hall, London.
- Higgins, S. I., Richardson, D. M., Cowling, R. M., and Trinder-Smith, T. H., 1999. Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. *Conservation Biology* 13, 303-313.
- Hosmer, D. W., and Lemeshow, S., 1989. *Applied Logistic Regression*. Wiley, New York.
- Huntley, B., Berry, P. M., Cramer, W., and McDonald, A. P., 1995. Modelling present and potential future ranges of some European higher plants using climate response. *Journal of Biogeography* 22, 967-1001.
- Lomolino, M. V., and Channell, R., 1998. Range collapse, re-introductions, and biogeographic guidelines for conservation. *Conservation Biology*, 481-484.
- Manel, D., Dias, J. M., Buckton, S. T., and Ormerod, S. J., 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology* 36, 734-747.
- Manel, S., Buckton, S. T., and Ormerod, S. J., 2000. Testing large-scale hypotheses using surveys: the effects of land use on the habitats, invertebrates and birds of Himalayan rivers. *Journal of Applied Ecology* 37, 756-770.
- Manel, S., Williams, H. C., and Ormerod, S. J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38, 921-931.
- McCullagh, P., and Nelder, J. A., 1989. *Generalized linear models*. Chapman & Hall,
- Midgley, G. F., Hannah, L., Millar, D., Thuiller, W., and Booth, A., In press. Developing regional and species-level assessments of climate change impacts on biodiversity: A preliminary study in the Cape Floristic Region. *Biological Conservation*.
- Monserud, R. A., and Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling* 62, 275-293.
- Pearce, J., and Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133, 225-245.
- Pearce, J., Ferrier, S., and Scotts, D., 2001. An evaluation of the predictive performance of the distributional models for flora and fauna in north-east New South Wales. *Journal of Environmental Management* 62, 171-184.
- Peterson, A. T., Sánchez-Cordero, V., Soberón, J., Bartley, J., Buddemeier, R. W., and Navarro-Sigüenza, A. G., 2001. Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecological Modelling* 144(1), 21-30.
- Peterson, A. T., and Vieglais, D. A., 2001. Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *BioScience* 51(5), 363-371.
- Rebelo, A. G., 1995. *Proteas, a field guide to the Proteas of southern Africa*. Fernwood Press, Cape Town.
- Rouget, M., Richardson, D. M., Cowling, R. M., and Lloyd, J. W., 2002. Current patterns of habitat transformation and future threats to biodiversity in terrestrial ecosystems of the Cape Floristic Region, South Africa. *Biological Conservation* In press.
- Rouget, M., Richardson, D. M., Milton, S. J., and Polakow, D., 2001. Predicting invasion dynamics of four alien *Pinus* species in a highly fragmented semi-arid shrubland in South-Africa. *Plant Ecology* 152, 79-92.
- Schulze, R. E., 1997. *South African Atlas of Agrohydrology and Climatology*, Water Research Commission, Pretoria.

- Spitz, F., and Lek, S., 1999. Environmental impact prediction using neural network modelling. An example in wildlife damage. *Journal of Applied Ecology* 36, 317-326.
- Stockwell, D. R. B., and Peterson, A. T., 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148, 1-13.
- Sykes, M. T., Prentice, I. C., and Cramer, W., 1996. A bioclimatic model for the potential distributions of north European tree species under present and future climate. *Journal of Biogeography* 23, 203-233.
- Thuiller, W., Araujo, M., and Lavorel, S., In review-a. Generalized Models versus Classification Tree Analysis: a comparative study for predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*.
- Thuiller, W., Vaydera, J., Pino, J., Sabaté, S., and Lavorel, S., In review-b. Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology & Biogeography*.
- Vayssières, M. P., Plant, R. E., and Allen-Diaz, B. H., 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* 11, 679-694.
- Yee, T. W., and Mitchell, N. D., 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science* 2, 587-602.

Table 1: Misclassification matrix summarizing the different types of errors provided by models. TN: true negative; TP: true positive; FP: false positive and FN: false negative.

		Predicted		Total
		Absence	Presence	
Observed	Absence	TN	FP	TN + FP
	Presence	FN	TP	FN + TP
Total		TN + FN	FP + TP	

Table 2. Index for classifying model prediction accuracy.

Accuracy	AUC	Kappa
Excellent or high	0.9 – 1	0.8 – 1
Good	0.8 – 0.9	0.6 – 0.8
Fair	0.7 – 0.8	0.4 – 0.6
Poor	0.6 – 0.7	0.2 – 0.4
Fail or null	0.5 – 0.6	0 – 0.2

Table 3: An example of misclassification matrix for the species *Aulax umbellata* (*Auumbe*) with a prevalence equal to 0.95. The sensitivity and specificity are derived from the misclassification matrix. Observed represents the count of the observed presences (symbolized by Pres) and absences (symbolized by Abs), Predicted represents the count of predicted presences (symbolized by Pres) and absences (symbolized by Abs). *se* = sensitivity and *sp* = specificity. AUC = area under the curve.

Prevalence = 0.95

		Threshold by ROC approach				Threshold by Kappa			
		Predicted				Predicted			
		Abs	Pres	Tot					
Observed	Abs	22	2	24	Observed	Abs	20	4	24
	Pres	31	432	463		Pres	0	463	463
	Tot	53	434			Tot	20	464	

True negative	22	20
Specificity	91.7%	83.3%
True positive	432	463
Sensitivity	93.3%	100%
False negative	31	0
1 - <i>se</i>	6.7%	0%
False positive	2	4
1 - <i>sp</i>	8.3%	16.7%
No of errors	33	4
Kappa	0.540	0.914
AUC	0.979	0.979

Table 4: An example of misclassification matrix for the species *Aulax umbellata* (*Auumbe*) with a prevalence equal to 0.05. The sensitivity and specificity are derived from the misclassification matrix. Observed represents the count of observed presences (symbolized by Pres) and absence (symbolized by Abs), Predicted represents the count of predicted presences (symbolized by Pres) and absences (symbolized by Abs). *se* = sensitivity and *sp* = specificity. AUC = area under the curve.

Prevalence = 0.05

Observed	Threshold by ROC approach				Observed	Threshold by Kappa			
	Predicted					Predicted			
	Abs	Pres	Tot			Abs	Pres	Tot	
Abs	8275	522	8797		Abs	8526	271	8797	
Pres	27	436	463		Pres	97	366	463	
Tot	8302	958			Tot	8623	637		

True negative	8275		8526
Specificity		94.1%	96.9%
True positive	436		366
Sensitivity		94.2%	79.1%
False negative	27		97
1 - <i>se</i>		5.8%	20.9%
False positive	522		271
1 - <i>sp</i>		5.9%	3.1%
No of errors	549		350
Kappa	0.586		0.645
AUC	0.992		0.992

FIGURE CAPTION

Figure 1: Three methods to construct and represent sensitivity, specificity and threshold using the ROC curve approach. a) Typical ROC curve graph, representing the sensitivity as a function of 1-specificity (false positive) calculated for various threshold values. AUC is calculated below this curve and above the straight line representing unity. b) Sensitivity as a function of specificity calculated for various threshold values. c) Sensitivity and specificity in function of various threshold values.

Figure 2: Overall predictive performance for threshold independent model evaluation methods. a) Graphic representing the evaluation of the area under the curve (AUC) for the 50 species as a function of prevalence. b) Graphic representing the evaluation of the adjusted-D² for the 50 species as a function of prevalence.

Figure 3: Specific predictive performance. The four boxplots represent the Kappa statistic calculated following the selected threshold in function of the species' prevalence including a) threshold = 0.5, b) threshold equals to the prevalence in the data, c) threshold determined by optimising Kappa, d) threshold from optimising by ROC.

Figure 4: Relation of sensitivity and specificity according to species prevalence, following the method to optimise the threshold. Sensitivity is labelled as (*se*) and specificity is labelled as (*sp*) on axes.

Figure 1

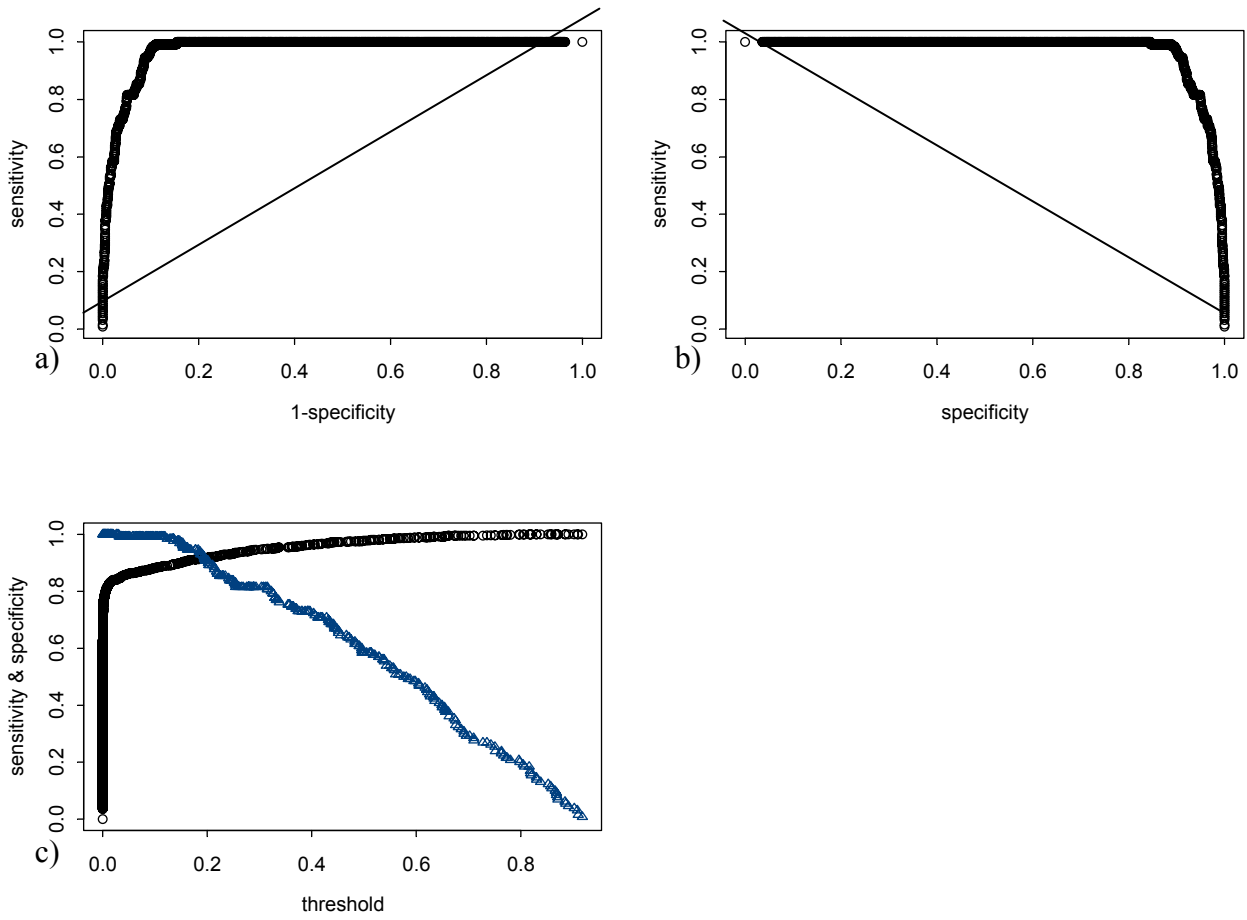


Figure 2

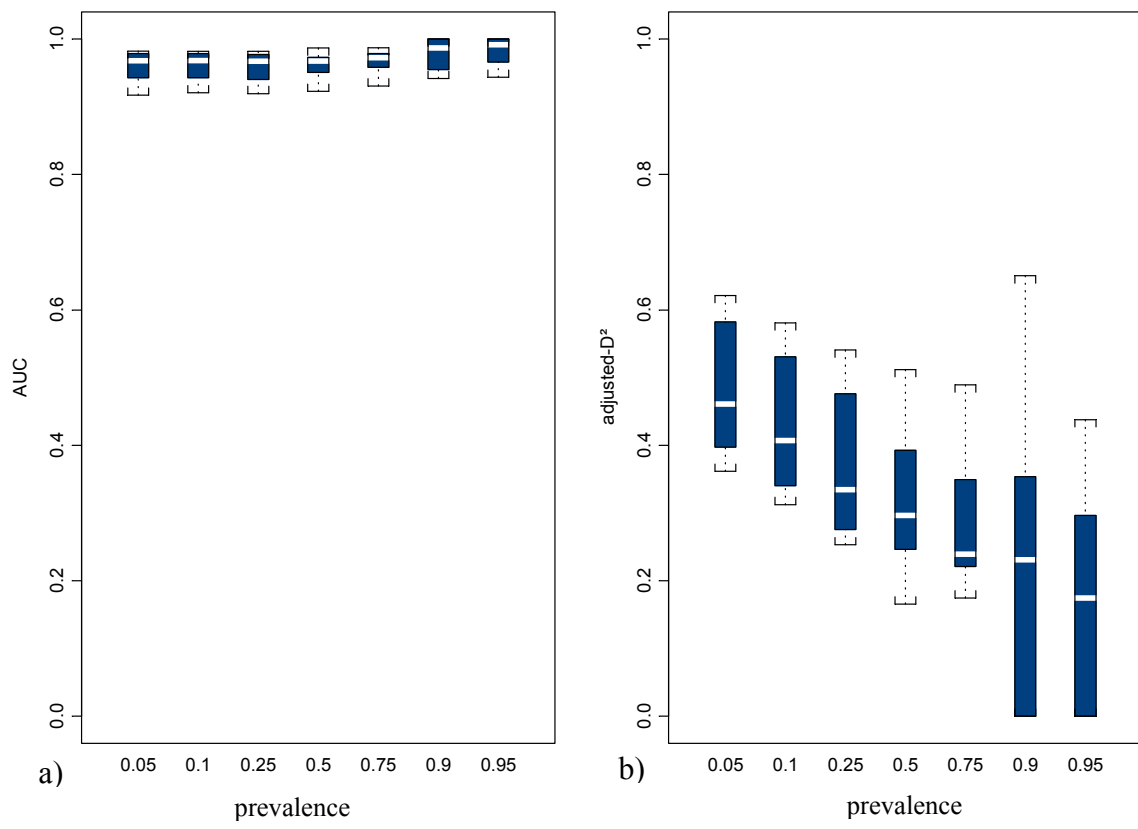


Figure 3

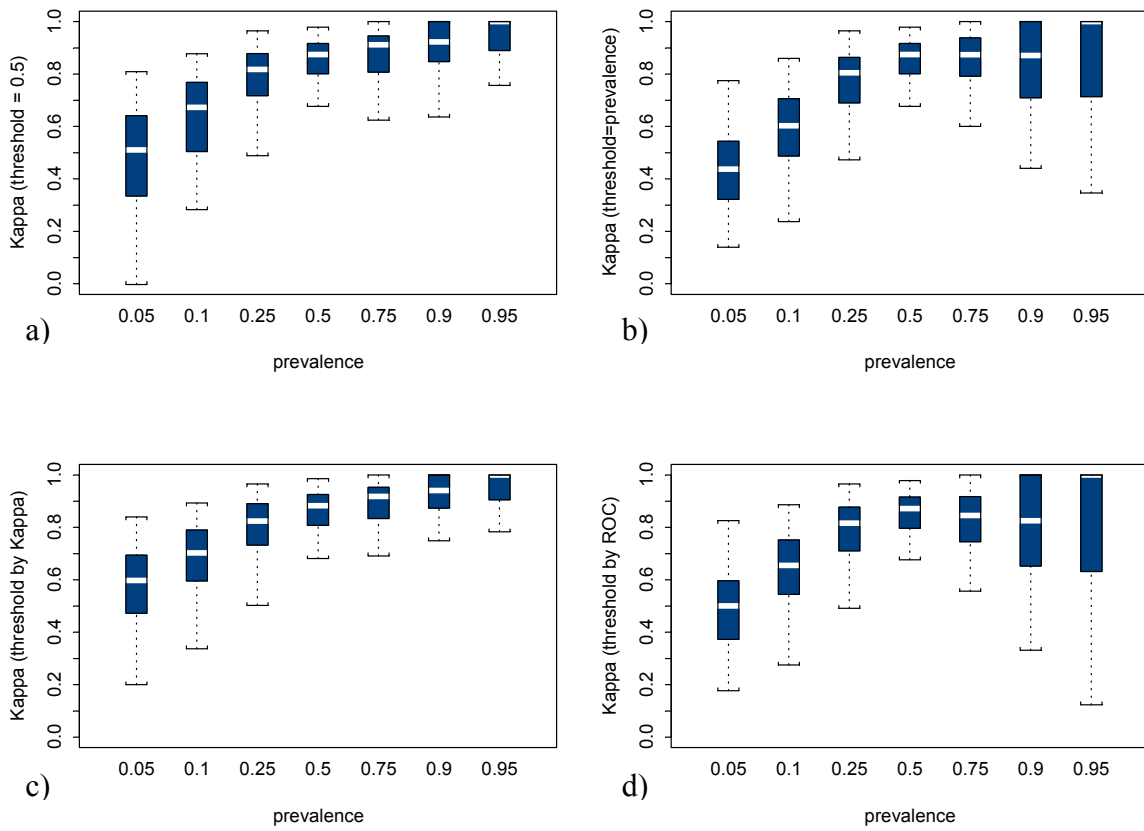
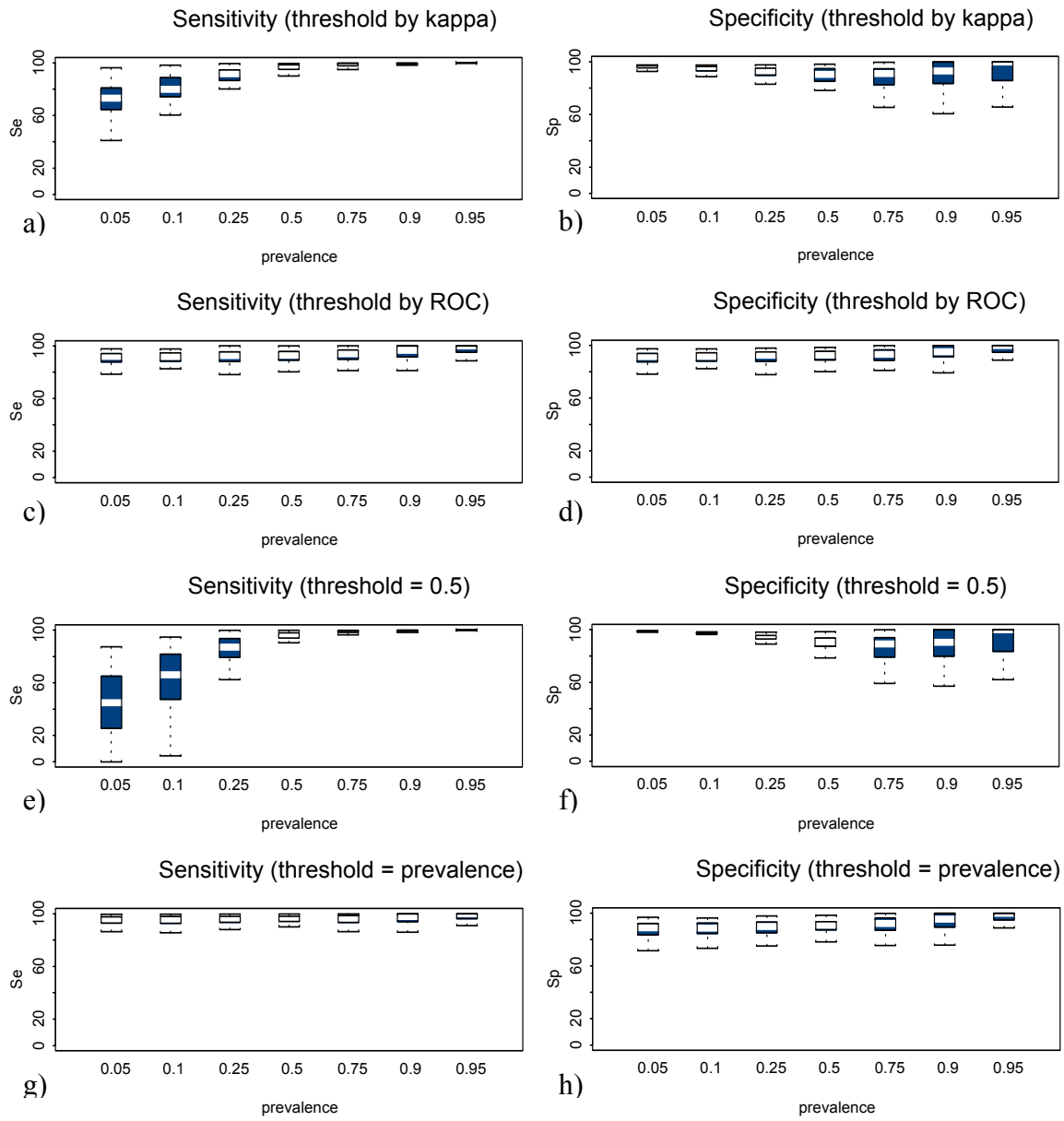


Figure 4



ANNEXE 1.3

Presence-absence vs presence-only modelling methods for predicting bird habitat suitability

Brotons L., Thuiller W., Araújo M.B., Hirzel A.H.

Ecography (2004) sous presse

PRESENCE-ABSENCE VERSUS PRESENCE-ONLY MODELLING METHODS FOR PREDICTING BIRD HABITAT SUITABILITY

Lluís Brotons^{1,2}, Wilfried Thuiller², Miguel Araújo^{2,3} and Alexandre Hirzel⁴

¹ Institut Català d'Ornitologia, C./ Girona 168, E-5, 08037 Barcelona, Spain.

² Corresponding address : Centre d'Ecologie Fonctionnelle et Evolutive-CNRS, 1919 Route de Mende, 34293 Montpellier, France. tel : (33) 467 613 423. E.mail : brotons@cefe.cnrs-mop.fr

³ Unidade de Macroecologia e Conservação, Universidade de Évora, 7000-730 Évora, Portugal

⁴ Institute of Zoology, Conservation Biology, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland.

Habitat suitability models can be generated using methods requiring information on species presence or species presence and absence. Knowledge of the predictive performance of such methods becomes a critical issue to establish their optimal scope of application for mapping current species distributions under different constraints. Here, we use breeding bird atlas data in Catalonia as a working example and attempt to analyse the relative performance of two methods. One using presence data only (Ecological Niche factor Analyses, ENFA) and the other using presence/absence data (Generalised linear models, GLM). Models were run on a set of forest species with similar habitat requirements, but with varying occurrence rate (prevalence) and ecological niche position (marginality). Our results support the idea that GLM predictions are more accurate than those obtained with ENFA when species use available habitats proportionally to their suitability, making absence data reliable and useful to enhance model calibration. Moreover, we showed that species marginality was strongly correlated to

predictive accuracy; with species having less restricted ecological requirements being modelled less accurately than more marginal ones, irrespective of the method employed. Models of more prevalent, tolerant species were more sensitive to the availability of absence data, suggesting that presence/absence methods may specially adequate for predicting distributions of this type of species. We conclude that model developers should keep in mind that species ecological characteristics are critical in determining the accuracy of the model and that it will be increasingly difficult to model generalist species independently of the methodology used. Being based on distinct approaches regarding adjustment to data and variation in data quality or availability, habitat distribution modelling methods cover different application areas, making it impossible to identify one among them as universally applicable. Our results suggest however, that if absence data is available, methods making use of its information should be preferably used in most situations.

INTRODUCTION

Mapping of species distributions is a key issue in ecology and conservation since the statement of hypotheses in these disciplines often relies on the accurate knowledge of where species occur in space. To map species distributions at large spatial scales, different approaches have been followed, the most common being the general atlas-like framework (Donald and Fuller 1998, Mitchel-Jones et al. 1999, Underhill and Gibbons 2002). The spatial positioning of data from large museum collections may also appear as an alternative in some cases (Peterson et al. 2002). Given the geographical extent of their coverage and the standardisation of their sampling methodology, atlases are however among the most powerful tools available to analyse species' distribution patterns and their governing factors (Donald and Fuller 1998). Nevertheless, most atlases have focused on reporting the occurrence of species and provide relatively poor quantitative information on species distribution and on

the relative suitability of different locations. Recent large scale atlas work (Gibbons et al. 1993) has attempted to obtain quantitative, fine grained data on species distributions to include quantitative variation in species distributions. Such quantitative atlases seek to go one step ahead by introducing some measure of species relative abundance or probability of occurrence in atlas cells (Johnson and Sargeant 2002).

Habitat suitability or niche based modelling techniques use information on species presence at different locations and environmental factors to generate statistical functions that allow quantitative predictions of potentially suitable habitat for a given species (for a review see Guisan and Zimmerman 2000). The projection of the generated functions to areas where environmental factors are known but species have not been sampled allows an optimal, cost effective method to quantitatively map species distributions in large regions but at low spatial resolutions (Guisan and Zimmerman 2000, Peterson et

al. 2002). The recent development of these techniques combined with increasing availability of large-scale environmental information in digital format offer an opportunity to test and improve methodologies for quantitative mapping of species distributions. Appropriate quantitative data on species distributions have already been demonstrated to provide useful information for conservation planning. For instance, species extinctions seem to be more likely in areas with low suitability, or areas where species are less abundant and including such information in reserve-selection improves the ability to ensure long term persistence of species (Araújo et al. 2002). Furthermore, habitat suitability models are increasingly being used to assess the impact of future land use or climate changes (Austin et al. 1996, Buckland et al. 1996, Peterson et al. 2002, Thuiller 2003), or design ecological networks at large spatial scales (Bani et al. 2002).

There are different methods available to generate habitat suitability maps for species. A major difference between them is the quality of data needed. A first group of methods includes generalised linear models (GLM), generalised additive models (GAM), tree regressions, and artificial neural networks (ANN). These methods require good quality presence/absence data in order to generate discriminative rules that allow suitability of habitats to be ranked according to the relative distributions of presences and absences in each of them (Manel et al. 1999, Guisan and Zimmerman 2000). On the other hand, methods such as ecological niche factors analysis (ENFA), Bioclim, or Domain, only require presence data. They have been developed to allow the use of data in which absences are of bad quality or unavailable (Carpenter et al. 1993, Hirzel et al. 2002 a, Farber and Kadmon 2003). These methods rely on the definition of environmental envelopes around locations where species occur, which are then compared to the environmental conditions of background areas (Hirzel et al. 2002a). Using a virtual species with predefined habitat selection preferences, Hirzel et al. (2001) compared model performances of a method relying on species presence only (ENFA) with a method that needs both presence and absence data (GLM). Although both methods provided good predictions of the virtual species distribution, these authors found that ENFA tended to outperform in hypothetical scenarios in which an invading species did not occupy all suitable habitat (i.e. many absences were thus detected in real habitat, spreading scenarios). Conversely, when species were using all optimal habitats with a high probability and sub-optimal habitats with lower probabilities (i.e. overabundance scenarios), GLM was more accurate.

Since data quality is likely to be a key issue affecting reliability of model predictions (Zaniewski et al. 2002, Stockwell and Peterson 2002 a,b), knowledge of the predictive performance of methods and their domain of application becomes a key issue at early stages of project development in surveys aimed at mapping species distributions. At present, we lack

extensive tests of the relative performances of methods that compare species distribution models using presence only or both presence/absence data. While the use of virtual species is useful as a preliminary exploration of the behaviour of methods, we need real data on species distributions to expand our understanding of the relative performances of methods. Also, since accurate data on absences is difficult to obtain, especially for mobile animal species, it is of especial interest to investigate when models using presence only data perform at least as well as models using presence/absence data. This should allow for a better understanding of the methods that are more adequate for particular applications. Another generally unexplored question is how dependent is the accuracy of a modelling approach to the ecological characteristics of the species and how these interact with species prevalence (i.e. proportion of occurrences in a data set) to determine model accuracy (but see Manel et al. 2001). Given that species with more restricted ecological niches are more localised and less frequent, one may predict, even in cases in which data quality is poor, that they are better modelled and thus their distributions more easily predicted than more widespread species (Stockwell and Peterson 2002b).

Here, we use breeding bird atlas data as a working example and attempt to analyse the relative performances of these two types of methods on a set of forest species with similar habitat requirements but with varying prevalence and ecological breadth. We first ask whether methods using presence data only perform equally well than those using presence/absence and whether hypothetical differences in performance hold when evaluating predictive habitat suitability methods on independent data sets. We then investigate the role of the ecological niche of species and prevalence on model accuracy and whether these factors interact with the method used to determine model accuracy..

METHODS

Bird data

In our assessment of habitat suitability methodologies, we used data on species occurrence as a subset of the Catalan Breeding Bird Atlas (CBBA, 1999-2002). The CBBA is a large-scale survey that covers the whole of the Catalan region (north-western Iberian Peninsula, 31.000 km², Fig. 1). Within the study area 1550 1x1 km cells were selected (covering approximately 6 % of the total area extent) to conduct standardised surveys of species presence during the breeding seasons 1999-2001. Cells were selected by volunteers in a stratified fashion assuming that they should cover representatively the main habitat types present in the 10x10 km Universal Transverse Mercator grid cell in which they were located (Hirzel and Guisan 2002). On each selected cell, two one-hour visits were conducted and the presence of species was detected. The first visit was made in March-April and the second during May-June to better cover the

breeding phenology span of different species. In this paper, we included species with at least 15 occurrences available and further selected species according to their prevalence so that the whole range of possible prevalence values were represented in the final data set (Table 1). Overall we analysed distributions of 30 forest species as judged from their habitat selection patterns in the Mediterranean area (Table 1).

Species presence in a given cell was assumed to be reliable. However, failing to detect a species does not mean that such species was truly absent from that cell. Presence is a probabilistic function mainly driven by species abundance and detectability. By assuming that detectability of a given species is constant across habitats, we considered that absences in this study were either reliable or associated to habitats in which abundance of species was low. Still, the assumption that absence indicates areas where the species are not present due to a negative species-environmental relationship is not necessarily a valid one. This assumption may not hold for a variety of reasons including habitat population dynamics, fragmentation, rate of dispersal or history, which may induce species absence from otherwise optimal habitat (Loehle and LeBlanc 1996). If the role of such events is significant and the species is not in equilibrium with its environment, absence data may not provide relevant information to model building and we expect methodologies relying on presence/absence data to perform worse than those relying on presence data only. On the other hand, if absences are indeed related to low suitable habitat for the species (i.e. the species is near the equilibrium with the environment), the information provided by them should improve the performance of methods relying on both presence and absence data (Hirzel et al. 2001).

Environmental data

Environmental variables (ENV) were generated from environmental information available as GIS (Geographical Information Systems) layers. Habitat composition was analysed from land-use rasters generated by the Cartographic Institute of Catalonia (ICC) and Agriculture Department (DARP, Table 2). After successive processes of simplification and classification, the definitive raster maps were resampled to 50 m pixels and converted to several boolean maps which allowed the generation of final variables describing each 1x1 km² cell (Table 2).

We also used climatic variables (temperature, precipitation and solar radiation) which were obtained from the Catalan Digital Atlas (CDA, Ninyerola et al. 2000), whereas data on relief was obtained from a Digital Elevation Model (DEM) generated by the ICC from topographic 1:50.000 maps. To obtain a value for each cell we calculated the mean value for all pixels (200 m side) in that cell (Table 2). We finally used three more variables that allowed us to detect geographic patterns in species distributions that were not captured by habitat or climatic ENV. These variables were the mean latitude and longitude co-

ordinate for each cell and the mean distance to the sea (Table 2).

Statistical models

Methods based on presence/absence data

Different methods have been envisaged to build up predictive models based on presence/absence data. Amongst them generalised linear models have been extensively tested elsewhere and have proved robust in a number of independent situations (Manel et al. 1999, Pearce and Ferrier 2000, Osborne et al. 2001, Thuiller et al. 2003). Therefore, to analyse binary data such as the presence/absence of species within each sampled cell, we applied generalised linear regression techniques with binomial error distribution (logistic regression, GLM, McCullagh and Nelder 1989). We included as potential predictors in model building all linear and quadratic terms for all the variables utilised, which excluded environmental predictors showing correlations higher than 0.9. To select the most parsimonious model, we used an automatic stepwise model selection procedure starting from a null model containing the intercept only. The "step.glm" function in S-Plus builds models by adding new terms and investigating how much they improve the fit, and by dropping terms that do not degrade the fit by a significant amount (Anon 1999). Quadratic terms were only present if they improved their linear counterpart. The statistic used to select the final model was the Akaike Information Criteria (AIC, Chambers and Hastie 1997). It is important to stress at this point that in this study GLM was used with a predictive rather than inductive objective in which accuracy of model predictions is more important than significance of particular ecological terms (Legendre & Legendre 1998).

Methods requiring presence data only

Different methods have been proposed to predict species distributions based on presence data only. These methods search for an "environmental envelope" characteristic of the points in which the species is present in order to extrapolate to the remaining area under study (Guisan and Zimmerman 2000). To analyse these kinds of data we used the Ecological Niche Factor Analysis (ENFA) released recently in the BIOMAPPER package (Hirzel et al. 2002b). ENFA quantifies the niche that a species occupies by comparing its distribution in ecological space ("the species distribution") with the distribution of all cells (the "global distribution") (Hirzel et al. 2002a). ENFA focuses on the marginality of the species (how the species mean differs from the global mean) and its specialization or environmental tolerance (how the species variance compares to the global variance). Species marginality gives indication of the species niche position whereas species tolerance rather refers to its niche width. ENFA uses a factor analysis with orthogonal rotations to (1) transform the predictor variables to a set of uncorrelated factors, and (2) to construct axes in a way that accounts for all marginality of the species in the first axis, and that

maximizes tolerance in the following axes. Still, different algorithms are available in BIOMAPPER package to build habitat suitability maps from ENFA analysis (Hirzel et al 2002). Following Hirzel and Arlettaz (2004) we used the geometric mean algorithm, which takes into account the density of observation points in environmental space by computing the geometric mean to all observation points. We used a Box-Cox transformation of the environmental variables to enhance normality except in the cases when transformation produced near binary outcomes (Hirzel et al. 2002a).

Marginal species are likely to be less tolerant in most conditions, and species marginality and tolerance were indeed highly correlated in our set of forest species ($r=-0.76$, $p<0.0001$). Furthermore, ecologically marginal species may tend to be less tolerant to changes in ecological conditions leading to restricted distributions. Species marginality and species prevalence were also significantly correlated in our data set ($r=-0.68$, $p<0.0001$). In order to allow the independent assessment of the different components of species niche we proceed to conduct a Principal Component Analysis using species marginality, tolerance and prevalence as original variables. After a varimax transformation of the principal components maximising their correspondence to the original variables, we succeed to obtain two independent components: (1) a marginality component strongly associated positively to species marginality ($r=0.90$) and more weakly, negatively to tolerance ($r=-0.60$), and (2) a prevalence component identifying a gradient of species prevalence ($r=0.90$) parallel to that of species tolerance ($r=-0.70$), separating less tolerant and scarcer species from more tolerant and abundant ones. These two components were used to investigate how niche characteristics affect predictive model accuracy.

Evaluation of habitat suitability models

We used cross-validation to evaluate predictive model accuracy and divided the data in two different sets by randomly assigning 70% of occurrence values for each species to a calibration data-set, and the 30% of remaining occurrences to an independent evaluation data set. The calibration data-set was used to develop the habitat model that was evaluated on the evaluation data set (Fielding and Bell 1997).

There are practical difficulties in evaluating predictions from presence only data models with traditional evaluation methods (Pearce and Ferrier 2000) given that absence data is usually missing and therefore can not be used to evaluate model predictions. A possible method is to compare the suitability of areas where the species is present with that of the background environment (Hirzel et al. 2001). Other authors, have used correlations with known, or reference distributions, to evaluate models performance (Hirzel et al. 2001, Zaniwski et al. 2002, Boyce et al. 2002). However, in our case true or reference species distributions were unavailable. Predicting species absences is an important issue even

when information has not been explicitly incorporated into model development (Stockwell and Peterson 2002a). Therefore, we assessed the accuracy of both ENFA and GLM models on the calibration and evaluation data sets using both presences and absences. By means of confusion matrices, results from probabilistic models are often judged as successful if predicted probabilities >0.5 correspond with observed occurrences and values <0.5 with absences and prediction errors (false positives and false negatives) are low. However, this dichotomy is arbitrary and lacks any ecological basis. A more powerful approach is to assess model success across a range of dichotomies from different cut-off points using receiver operating characteristics (ROC) plots. The ROC plot is based on a series of confusion matrices computed for a range of cut-offs from 0 to 1. It then plots on the y-axis the true positive fraction, against the false positive fraction from the same confusion matrix (Fielding and Bell 1997, Pearce and Ferrier 2000). The area under the ROC curve (AUC) is a convenient measure of overall fit and commonly varies between 0.5 (for chance performance) and 1 (perfect fit). We obtained AUC and its standard error with a custom function in S-Plus software (Anon 1999).

Comparison of accuracy between modelling methodologies

We first test for overall differences between modelling method (GLM vs. ENFA) and data-set (calibration vs. evaluation) by means of repeated measures ANOVA using modelling method and data-set as within-subject factors in the design according to species. We then used repeated measures ANCOVA designs to assess how accuracy of habitat models varied between method and data-set using these factors as within-factors subjects but also adding to the design the two covariables related to niche characteristics and species prevalence (i.e. the marginality component and the prevalence component).

RESULTS

Overall accuracies of models

Overall model accuracy estimated via the ROC method performed better than random in every case analysed (Table 1). AUC values were higher for GLM models than for ENFA models (Table 3, Fig. 2 and 3) and were also higher when evaluated for the calibration data compared to the evaluation data set (Table 3, Fig. 2). We also found that change in predictive accuracies between the calibration and the evaluation data sets was larger for GLM than for ENFA models, indicating that the loss in predictive performance when applied to an independent data set not used for model construction is higher for GLM (Table 3, Fig. 2). In addition to overall differences in predictive accuracy, we detected considerable inconsistency in the spatial projections of GLM and ENFA models (Fig 1). There was a general tendency for ENFA to overestimate the spatial extent of the

distributions over the edges of those estimated by GLM, whereas in some cases, high quality areas estimated by GLM were overlooked by ENFA (e.g. Fig. 1).

Effects of species niche characteristics and prevalence on model performance

Species ecology had a major role in the determination of species model accuracies, with AUC being generally higher and thus, distribution better predicted in more marginal species (marginality component, Table 3). The positive effect of marginality on predictive model accuracy was similar among the different methods and data sets as indicated by the lack of significant interactions suggesting that this effect was highly robust to methodological considerations (Table 3, Fig. 4).

The prevalence component did not have an overall consistent effect on model performance (Table 3, Fig. 5). However, there were significant differences in the effect of this factor on model performance between methods in the evaluation data set but not in the calibration data set, with higher values of the prevalence component associated with higher AUC values in GLM models but not in ENFA which remained unaffected by this factor (Table 3, Fig. 5). This effect resulted in a stronger overall positive relationship between the prevalence component and predictive accuracy in the evaluation data set than in the calibration data set (Table 3, Fig. 5).

DISCUSSION

Our results showed that GLM using both presence and absence data predicted the distribution of forest species with higher accuracy than ENFA, which used presence data only. This supports the view that the forest species analysed used available habitats proportionally to their suitability, making absence data reliable and useful to enhance model calibration. This is in line with the results obtained by Hirzel et al. (2001) using a modelling approach based on a virtual species with predetermined habitat preferences. The authors found that GLM performed significantly better than ENFA when estimating habitat suitability in an overabundance scenario in which species occupied all optimal habitats and occupied secondary habitats at lower probabilities. In this scenario, absence data is likely to be reliable and help to “fix the floor” of what is unsuitable habitat for each focal species. First, by giving a low weighting to occurrences in low density habitats, absence data helps to identify low suitability areas that may have otherwise been classified as good habitats if only presence data were used. Occurrences in good but scarce habitats may also bias models based on presence only data because relative importance of such habitats may be over-weighted by a larger number of observations in other habitat types. For instance in the case of the nuthatch, deciduous forest areas, which cover a limited surface within the study region, were ranked as low suitability by the presence only method. Indeed this species had a small number of occurrences in such areas which were overridden by

the greater number of occurrences in other more abundant habitats (Fig 1). In such cases the availability of absences may become critical to correctly assess the relative suitability of these areas in comparison with other areas equally suitable but where presences are more common due to the relative availability of different habitats in the area. Some authors have suggested that in the lack of absence data, distribution models may be improved by generating random pseudo-absences from background areas in which species data is missing (Stockwell and Peterson 2002a). However, this method may result in biased absence data if species are widespread or presence data is scarce (Boyce et al. 2002).

The approach employed to test model accuracy assumes that the identification and prediction of the locations from which species is absent is of interest. This is likely to be the case in species based, cartographic habitat modelling works such as vertebrate atlas studies, but may not be the case if the objectives are more specific. In cases in which assumptions on species equilibrium are strongly violated (i.e. use of museum data to produce distributions at large spatial scales or distribution of invading or spreading species in which absence data is ecologically meaningless), evaluation of absence data becomes meaningless. Presence only methods may then be most useful and make best use of available presence data. They also have the advantage of reducing the contributions of non-equilibrium factors in model predictions (Hirzel et al. 2002a). Recently, Zaniewski et al. (2002), have shown that although presence-absence based methods were more discriminant than ENFA at a species level, they appeared less suitable to identify areas with high conservation concern from a multi-specific perspective. In this case, if the objective is to protect rare or endangered species, overestimating areas of potentially elevated biodiversity might be more preferable to underestimating their existence and presence only methods may be useful (Zaniewski et al. 2002).

Sudden changes in habitat quality may occur in nature situations (Gates and Donald 2000) resulting in individuals not using optimal habitats or being present in low quality areas. Caution in the use of habitat suitability methods should be advised if strong suspicions of non-equilibrium situations are expected. However, in large scale distribution modelling, most species, especially in rather mobile groups such as birds, are likely to be close to equilibrium with environmental conditions due to population dynamics and habitat selection mechanisms (Chamberlain and Fuller 1999). In these cases, absences are likely to reflect low habitat suitability and therefore improve model performance. We therefore argue that the use of presence/absence data is generally more appropriate than presence only data to build habitat suitability models and quantitative distributions of species in intensive data collection studies such as breeding bird atlases. In these approaches, quantitative spatial

distribution determined by the species realised niche is likely to prove highly useful. Ecological interpretation of different habitat modelling methodologies is of great importance and may guide the final choice of available alternatives. Zaniwski et al. (2002) argue that pure presence-only methods such as ENFA are more likely to predict potential distributions that more closely resemble the fundamental niche of the species, whereas presence-absence modelling is more likely to reflect the present natural distribution derived from realized niche. However, both methods aim at predicting distributions by sampling real distributions, and therefore, they provide different estimations of the realised niche of the species (Loehle and Leblanc 1996). Since presence only methods do not take into account the areas from which the species might be absent, they are simply less conservative in estimating the species realised niche. On the other hand, they may better capture realised niche responses in species which are far from equilibrium with the environment and therefore are not yet using all habitats corresponding to their realized niche (Hirzel et al. 2001).

It is important to emphasize that, being based on distinct approaches regarding adjustment to data and variation in data quality, habitat distribution modelling methods will likely cover different application areas and it will be impossible to identify one among them as universally applicable (Elith and Burgman 2002 a, b). Therefore, the goals and assumptions of habitat modelling should be clear before they are applied to particular situations. Methods based on presence only data such as ENFA appear to fully cover habitat modelling focused on data in which absence data is not available, or when the main objective of the modelling is to identify overall suitable areas for a given species in cases in which the current distribution of the species is known to be unreliable. Otherwise, methodologies employing presence/absence methods should be prioritised.

Species niche characteristics, prevalence and model accuracy

We found that ecological niche position (marginality) plays a key role in determining predictive accuracy in models developed with both GLM and ENFA methodologies. In particular, less marginal bird species from which selected habitats differed little from the available environmental conditions in the study area were modelled less accurately than more marginal selective species. This result agrees with the results of Hepinstall et al. (2002) and Stockwell and Peterson (2002b) who also observed that the performance of bird habitat models was negatively correlated with the proportion of habitats used by a species with more generalist species being poorly modelled. Stockwell and Peterson (2002b) offered as a biological explanation for this observation that widespread species often show local or regional differences in ecological characteristics. Modelling all these sub-populations together would effectively overestimate the species ecological breath

and hence reduce model accuracy. Therefore, the more widespread a species is, the more likely it is to use different habitats thus increasing the likelihood that more factors determine its distributions (Osborne and Seoane 2002). In such cases, accurate prediction of species distribution becomes difficult and will benefit from the availability of absence data to determine relative suitability among available habitats. By contrast, both GLM and ENFA methods seem to perform equally well on more marginal species, which offers a promising background to the development of models of marginal potentially threatened species from sources of poor quality data (Peterson et al. 2002).

A major but variable role of prevalence on the predictive accuracy of habitat models has been stressed by several works (Pearce and Ferrier 2000, Manel et al. 2001, Stockwell and Peterson 2002b). Manel et al. (2001) found that predictive model accuracy judged by the ROC method was independent of prevalence. However, a critical assessment of effects of prevalence on model predictive accuracy is problematic because prevalence is likely to vary both with species ecological character and relative sampling effort. More marginal or less tolerant species will tend to be less frequent and therefore, relatively fewer occurrences will be available than for species with a wider ecological spectra. On the other hand, relative lower sampling effort or bias in data collection may also decrease species prevalence. Prevalence is thus likely to affect model accuracy more strongly via indirect effects of species ecology. Thuiller et al. (in review) found that within a given species, accuracy is independent of prevalence supporting the view that among species effects of prevalence on model accuracy are likely to be associated to variability in species niche characteristics. In our study, we could not completely isolate prevalence from this factor. However, we found that independently of the marginality component, the effect of the prevalence component may still play a secondary role on predictive model accuracy (Hirzel et al. 2001, Karl et al. 2002). This role suggests that the effect of prevalence on predictive accuracy is moderately stronger in models using presence/absence data, because a relative increase in the amount of information derived from the additional presences may enhance its ability to discriminate the quality of the different sites. When using a presence only method an increase in the number of occurrences analysed did not render similar benefits to model accuracies. Indeed, using a virtual species, Hirzel et al. (2001) already showed that independently of data quality ENFA appeared to be robust to data quantity. This is supported by our results that the prevalence component did not affect accuracies of ENFA models independently from species ecology. In Hirzel et al. (2001), GLM was also found to be relatively robust to data quantity. However, our results suggested that higher prevalence for a given species ecology may enhance model accuracy on independent test data, raising the issue of the importance of testing habitat

predictive models on evaluation data tests not used for model development (Fielding and Bell 1997, Beutel et al. 1999, Hirzel et al. 2001). Further, studies should explicitly assess the influence of the relationship between sample size and ecology on the relative performance of habitat suitability models based on presence and absence methods.

ACKNOWLEDGEMENTS

We thank all volunteers that contributed to the collection of data for the Catalan Breeding Bird Atlas (CBBA), especially the two coordinators of the project, Joan Estrada and Vittorio Pedrocchi for their effort on data collection and management. The CBBA was financed by the "Departament de Medi Ambient de la Generalitat de Catalunya", "Fundació Territori i Paisatge" and "Sociedad Española de Ornitología". We thank G. Sirwardena, M. Mönkkönen and anonymous referee for improving earlier drafts of the manuscripts. This research is a contribution to the Montpellier-Barcelona LEA "Mediterranean Ecosystems in a Changing World" and has been supported by a Marie Curie Fellowship of the European Community programme Improving Human Potential under the contract number HPMF-CT-2002-01987.

REFERENCES

- Anon 1999., S-Plus 2000 Guide to Statistics. Volume 1. Seattle, Mathsoft.
- Araujo, M.B., Williams P.H. and Fuller, R.J., 2002. Dynamics of extinction and the selection of nature reserves. - *Proceedings Royal Society, Biological Sciences*, 269:1971-1980.
- Austin, G. E., Thomas, C.J., Houston, D.C. and Thompson, D.B.A., 1996. Predicting the spatial distribution of buzzard *Buteo buteo* nesting areas using a Geographical Information System and remote sensing. - *Journal of Applied Ecology*, 33:1541-1550.
- Bani, L., Baietto, M., Bottoni, L and Massa, R., 2002. The use of focal species in designing a habitat network for a lowland area of Lombardy, Italy. - *Conservation Biology*, 16:826-831.
- Beutel, T.S., Beeton, R.J.S. and Baxter, G.S., 1999. Building better wildlife-habitat models. - *Ecography*, 22:219-223.
- Boyce, M., Vernier, P.R., Nielsen, S. and Schmiegelow, F.K. 2002., Evaluating resource selection functions. - *Ecological Modelling*, 157, 281-300.
- Buckland, S. T., Elston, D. A., and Beaney, S. J., 1996. Predicting distributional change, with application to bird distributions in northeast Scotland. - *Global Ecology and Biogeography Letters*, 5, 66-84.
- Carpenter, G., Gillison, A.N. and Winter, J., 1993. Domain: a flexible modelling procedure for mapping potential distributions of plants and animals. - *Biodiversity and Conservation*, 2: 667-680.
- Chamberlain, D.E. and Fuller, R.J., 1999. Density-dependent habitat distribution in birds: issues of scale, habitat definition and habitat availability. - *Journal of Avian Biology*, 30: 427-436.
- Chamber J.M. and Hastie T.J., 1997. Statistical models in S. - Chapman and Hall.
- Donald, P.F. and Fuller, R., 1998. Ornithological atlas data: A review of uses and limitations. - *Bird Study*, 45: 129-145.
- Elith, J, and Burgman, M.A., 2002. Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. In: *Predicting species occurrences: issues of accuracy and scale*. Scott, M.S. et al. Island Press. Washington.
- Elith, J., and Burgman, M. A., 2002. Habitat models for PVA. In: *Population Viability in Plants*, Editors: Brigham C.A. and Schwartz M.W. Springer-Verlag, New York. In press.
- Farber, O. and Kadmon, R., 2003. Assessment of alternative approaches for bioclimatic modelling with special emphasis on the Mahalanobis distance. - *Ecological Modelling*, 160: 115-130.
- Fielding A.H. and Bell J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. - *Environmental Conservation* 24:38-49.
- Gates, S., and Donald, P. F. 2000. Local extinction of British farmland birds and the prediction of further loss. - *Journal of Applied Ecology*, 37:806-820.
- Gibbons, D.W., Reid, J. B. and Chapman, R. A., 1993. *The new Atlas of Breeding Birds in Britain and Ireland: 1988-1991*. T and AD Poyser. London
- Guisan A. and Zimmerman N.E., 2000. Predictive habitat distribution models in ecology. - *Ecological Modelling*, 135:147-186.
- Hausser, J., 1995. *Mammifères de Suisse*. Bâles, Birkhäuser.
- Hepinstall, J.A., Krohn, W.B. and Sader, S.A., 2002. Effects of niche width on the performance and agreement of avian habitat models. In: *Predicting species occurrences: issues of accuracy and scale*. Scott, M.S. et al. Island Press. Washington.
- Hirzel, A. H., and Arlettaz, R. 2004. Modelling habitat suitability for complex species distributions by the environmental-distance geometric mean. - *Environmental Management*, in press.
- Hirzel, A.H., and Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modelling. - *Ecological Modelling*, 157, 331-341.
- Hirzel, A.H., Helfer, V. and Métral, F., 2001. Assessing habitat-suitability models with a virtual species. - *Ecological Modelling*, 145, 111-121.
- Hirzel A.H., Hausser J., Chessel D. and Perrin N., 2002. Ecological-niche factor analysis: How to compute habitat- suitability maps without absence data? - *Ecology*, 83, 2027-2036
- Hirzel, A.H., Hausser, J. and Perrin N., 2002a. *Biomapper 2.0*. Division of Conservation Biology, Bern.
- Johnson, D.H. and Sargeant, G.A., 2002. Towards better atlases: Improving Presence-Absence Information. In: *Predicting species occurrences: issues*

- of accuracy and scale. Scott, M.S. et al. Island Press. Washington.
- Karl, J.W., Bomar, L.K., Heglund, P.J., Wright, N.M. and Scott, M., 2002. Species commonness and the accuracy of habitat-relationship models. In: Predicting species occurrences: issues of accuracy and scale. Scott, M.S. et al. Island Press. Washington.
- Loehle, C., and LeBlanc, D., 1996. Model-based assessments of climate change effects on forests: a critical review. - *Ecological Modelling*, 90: 1-31.
- Manel, D., Dias, J. M., Buckton, S. T., and Ormerod, S. J., 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. - *Journal of Applied Ecology*, 36, 734-747.
- Manel, S. Williams, H.C. and Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. - *Journal of Applied Ecology*, 38:921-931.
- McCullagh, P. and Nelder, J.A. 1989. *Generalized Linear Models*, London: Chapman Hall.
- Mitchell-Jones, A. J., Amori, G., Bogdanowicz, W., Krystufek, B., Reijnders, P. J. H., Spitzenberger, F., Stubbe, M., Thissen, J. B. M., Vohralik, V. and Zima, J. (red.), 1999. Atlas of European Mammals. Academic Press, London. 496 s.
- Ninyerola M., Pons X. and Roure J.M., 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. - *International Journal of Climatology*, 20:1823-1841.
- Osborne P.E., Alonso J.C. and Bryant R.G., 2001. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. - *Journal of Applied Ecology*, 38:458-471.
- Osborne, P.E. and Suárez-Seoane, S., 2002. Should data be portioned before building large-scale distribution models? - *Ecological Modelling*, in press.
- Pearce, J. and Ferrier, S., 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. - *Ecological Modelling*, 128:127-147.
- Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sánchez-Cordero, V., Soberón, J. Buddemeier, R.H. and Stockwell, D.R., 2002. Future projections for Mexican faunas under global change scenarios. - *Nature*, 416:626-629.
- Stockwell, D.R. and Peterson, A.T., 2002a. Controlling bias in biodiversity. In: Predicting species occurrences: issues of accuracy and scale. Scott, M.S. et al. Island Press. Washington.
- Stockwell, D.R. and Peterson, A.T., 2002b. Effects of sample size on accuracy of species distribution models. - *Ecological Modelling*, 148:1-13.
- Thuiller, W. (2003) BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9, 1353-1362.
- Thuiller, W. 2003. Impacts of global climate change on biodiversity in Europe: projections and uncertainties. Thesis. University of Montpellier, Montpellier.
- Thuiller, W., Araújo M.B., Lavorel S. (2003) Generalized Models versus Classification Tree Analysis: a comparative study for predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* 14, 669-680.
- Thuiller, W., Brotons L., Araújo M.B., Lavorel S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography in press*.
- Underhill, L. and Gibbons, D., 2002. Mapping and monitoring bird populations: their conservation uses. In: *Conserving Bird Biodiversity: General Principles and Applications*, Norris, K. and Pain, D.J. (eds.). University Press. Cambridge.
- Zaniewski, A.E., Lehman, A. and Overton, J., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. - *Ecological Modelling*, 157, 261-280..

Table 1. Species and model accuracies \pm S.E. as estimated from area under the ROC curve (AUC). § Marginality as estimated by Biomapper algorithm. Species are sorted according to their prevalence which was calculated as proportion of presences within the data set of 1550 cells sampled.

Species names	GLM (AUC)		ENFA (AUC)		Prevalence	Marginality Index§
	Calibration	Evaluation	Calibration	Evaluation		
<i>Emberiza citrinella</i>	0.95 \pm 0.036	0.93 \pm 0.041	0.88 \pm 0.015	0.95 \pm 0.024	0.01	0.73
<i>Parus palustris</i>	1.00 \pm 0.002	0.94 \pm 0.028	0.82 \pm 0.045	0.91 \pm 0.013	0.01	0.93
<i>Phoenicuros phoenicuros</i>	0.91 \pm 0.027	0.57 \pm 0.182	0.67 \pm 0.012	0.66 \pm 0.111	0.02	0.65
<i>Regulus regulus</i>	0.99 \pm 0.003	0.91 \pm 0.044	0.94 \pm 0.043	0.91 \pm 0.091	0.03	0.99
<i>Sylvia hortensis</i>	0.89 \pm 0.021	0.76 \pm 0.069	0.72 \pm 0.004	0.68 \pm 0.056	0.03	0.59
<i>Anthus trivialis</i>	0.90 \pm 0.024	0.79 \pm 0.066	0.80 \pm 0.028	0.76 \pm 0.071	0.04	0.69
<i>Prunella modularis</i>	1.00 \pm 0.001	0.91 \pm 0.001	0.94 \pm 0.019	0.89 \pm 0.053	0.05	0.88
<i>Sylvia borin</i>	0.82 \pm 0.022	0.79 \pm 0.043	0.71 \pm 0.028	0.72 \pm 0.028	0.06	0.66
<i>Sitta europaea</i>	0.93 \pm 0.010	0.91 \pm 0.016	0.82 \pm 0.023	0.81 \pm 0.037	0.11	0.77
<i>Muscicapa striata</i>	0.76 \pm 0.022	0.68 \pm 0.040	0.58 \pm 0.014	0.60 \pm 0.019	0.12	0.62
<i>Emberiza cia</i>	0.88 \pm 0.011	0.86 \pm 0.019	0.76 \pm 0.025	0.75 \pm 0.041	0.21	0.55
<i>Turdus philomelos</i>	0.81 \pm 0.015	0.78 \pm 0.022	0.70 \pm 0.016	0.71 \pm 0.024	0.25	0.57
<i>Phylloscopus collybita</i>	0.84 \pm 0.014	0.75 \pm 0.027	0.69 \pm 0.017	0.64 \pm 0.026	0.25	0.62
<i>Parus ater</i>	0.92 \pm 0.009	0.89 \pm 0.016	0.67 \pm 0.017	0.64 \pm 0.027	0.27	0.66
<i>Sylvia cantillans</i>	0.83 \pm 0.013	0.81 \pm 0.020	0.69 \pm 0.017	0.68 \pm 0.025	0.29	0.47
<i>Oriolus oriolus</i>	0.77 \pm 0.015	0.68 \pm 0.026	0.66 \pm 0.013	0.62 \pm 0.021	0.29	0.54
<i>Lullula arborea</i>	0.87 \pm 0.011	0.78 \pm 0.021	0.73 \pm 0.017	0.70 \pm 0.027	0.30	0.47
<i>Regulus ignicapillus</i>	0.88 \pm 0.010	0.84 \pm 0.018	0.81 \pm 0.016	0.76 \pm 0.025	0.35	0.60
<i>Parus cristatus</i>	0.83 \pm 0.012	0.84 \pm 0.018	0.74 \pm 0.013	0.76 \pm 0.022	0.37	0.52
<i>Streptotelia turtur</i>	0.81 \pm 0.013	0.81 \pm 0.020	0.72 \pm 0.014	0.69 \pm 0.021	0.41	0.56
<i>Aegithalos caudatus</i>	0.86 \pm 0.011	0.83 \pm 0.019	0.78 \pm 0.015	0.81 \pm 0.023	0.38	0.57
<i>Phylloscopus bonelli</i>	0.84 \pm 0.012	0.79 \pm 0.021	0.69 \pm 0.016	0.69 \pm 0.025	0.43	0.47
<i>Troglodytes troglodytes</i>	0.87 \pm 0.010	0.80 \pm 0.020	0.77 \pm 0.014	0.72 \pm 0.023	0.48	0.54
<i>Sylvia melanocephala</i>	0.93 \pm 0.007	0.90 \pm 0.014	0.75 \pm 0.015	0.75 \pm 0.022	0.50	0.63
<i>Parus caeruleus</i>	0.88 \pm 0.010	0.82 \pm 0.020	0.77 \pm 0.014	0.76 \pm 0.023	0.51	0.52
<i>Garrulus glandarius</i>	0.85 \pm 0.011	0.81 \pm 0.020	0.76 \pm 0.015	0.72 \pm 0.025	0.55	0.51
<i>Emberiza cirrus</i>	0.84 \pm 0.012	0.81 \pm 0.020	0.75 \pm 0.015	0.73 \pm 0.024	0.55	0.47
<i>Eritacus rubecula</i>	0.91 \pm 0.009	0.90 \pm 0.014	0.83 \pm 0.013	0.83 \pm 0.020	0.57	0.53
<i>Luscinia megharrynchos</i>	0.83 \pm 0.014	0.80 \pm 0.023	0.70 \pm 0.018	0.69 \pm 0.027	0.66	0.51
<i>Turdus merula</i>	0.88 \pm 0.015	0.85 \pm 0.022	0.60 \pm 0.016	0.64 \pm 0.024	0.85	0.41

Table 2. Environmental variables (ENV) used to generate habitat suitability models of the 30 forest bird species used in the method comparison. Unless otherwise mentioned, variables referred to 1x1 km squares correspond to means obtained from averaging individual values from pixels contained each 1x1 km square. Cartographic sources are indicated when necessary.

Descriptor type	Variable description	Range of values
Forest	Coniferous forest ¹	0-400
	Esclerophylous ¹	0-400
	Deciduous forest ¹	0-400
	<i>Pinus halepensis</i> forest ²	0-400
	<i>Pinus sylvestris</i> forest ²	0-400
	<i>Abies alba</i> - <i>Pinus uncinata</i> forest pixels in 1x1 km squares ²	0-400
	<i>Pinus nigra</i> forest ²	0-400
	Other <i>Pinus</i> forest ²	0-400
	<i>Quercus suber</i> forest ²	0-400
	<i>Quercus ilex</i> forest ²	0-400
	<i>Quercus humilis</i> forest ²	0-393
	Other deciduous forest ²	0-400
	Distance to nearest forest patch (log m) ²	0-10
Agriculture	Dry herbaceous cropland (cereals) ¹	0-400
	Irrigated herbaceous cropland pixels (corn) ¹	0-400
	Dry arboreal cropland (olive tree, almond) ¹	0-400
	Irrigated arboreal cropland (fruit trees) ¹	0-400
	Vineyard ¹	0-400
Low vegetation cover	Scrub ¹	0-400
	Bare ground (rocks) ¹	0-400
Landscape	Number of land uses in 1x1 km squares (based on land use cover 1997, urban and industrial categories clumped) ¹	1-11
Human impact	Low density urbanization ¹	0-190
	Distance to cities >10.000 h (log m) ²	0-11
	Infrastructure (transport network and urban areas) ¹	0-400
	Distance to main roads of the primary road network (log m) ¹	0-10
	Distance to roads of the secondary road network (log m) ¹	0-10
Climate	Mean solar radiation ³ (10kJm ⁻² ·día ⁻¹)	19-961
	Mean accumulated summer precipitation (June-September) (l/m ²) ³	50-500
	Mean accumulated of mean winter temperatures (December-March) (degrees) ³	-50-105
Relief	Mean altitude (m) ¹	0-2850
	Mean slope ¹	0-39
	Mean latitude ¹	2.70-3.80
Others	Mean longitude ¹	45.70-46.50
	Mean distance to the sea (km) ¹	33-100

¹ 'Institut Cartogràfic de Catalunya' (ICC)

² 'Departament de Medi Ambient de la Generalitat de Catalunya' (DAM)

³ 'Centre de Recerca Ecologica i Aplicacions Forestals' (CREAF)

Table 3. Repeated-measures ANCOVA conducted on the predictive model accuracies of GLM and ENFA models on 30 forest species in Catalonia. The within subject effects considered are method (two levels, GLM vs. ENFA, see section methods) and data-set (two levels, calibration vs. evaluation). Species marginality and relative prevalence (see methods) were used as continuous predictors in the ANCOVA analyses. Significant results are emphasized in bold.

Source of variation	Model accuracy (AUC)		
	d.f.	F	P
	Between subject effects		
Marginality	1	27.23	<0.0001
Relative prevalence	1	6.88	<0.01
Error	27		
	Within subject effects		
Method	1	38.64	<0.0001
Method*Marginality	1	9.50	<0.01
Method*Relative prevalence	1	1.84	0.19
Error	27		
Data-set	1	0.20	0.53
Data-set*Marginality	1	0.40	0.16
Data-set*Relative prevalence	1	3.58	0.07
Error	27		
Method*Data-set	1	0.12	0.74
Method*Data-set*Marginality	1	1.90	0.18
Method*Data-set*Relative prevalence	1	4.44	<0.05
Error	27		

Figure legends

Fig. 1. Maps showing the predicted distributions of GLM (A) and ENFA (B) and the discrepancies between the two methods (C) for one of the forest species used, the Nuthatch (*Sitta europaea*). The discrepancy map was created by crossing predicted binary presence/absence maps after choosing for each modelling method a threshold maximising specificity and sensitivity. In (C) black cells show areas where ENFA predicted species absence and GLM presence, whereas light grey cells show areas where ENFA predicted presence and GLM predicted absence. Dark grey indicate coincidence in model predictions.

Fig. 2. Mean values of model accuracies for GLM and ENFA methods applied on both the calibration (70 % of sample cells, triangles) and the evaluation data (remaining 30% of sample cells, circles, n=30 species). Whiskers represent the standard error of the estimates.

Fig. 3. ROC plot for GLM and ENFA methods for the evaluation data set on one of the species analysed, the nuthatch, *Sitta europaea* (see figure 1 for predicted distributions, and table 1 for actual ROC values). Sensitivity represents the true positive fraction and 1-specificity the false positives fraction for a range of cut-offs used to classified modelled probabilities into presence absence data.

Fig. 4. Relationship between model accuracies for GLM and ENFA methods and species marginality component (index describing how far from the mean conditions of the study area the species optimum is found, n=30 species). A. Calibration data (70 % of sample cells). B. Evaluation data (remaining 30% of sample cells).

Fig. 5. Relationship between model accuracies for GLM and ENFA methods and the prevalence component (gradient separation abundant tolerant species from scarce less tolerant species, n=30 species). A. Calibration data (70 % of sample cells). B. Evaluation data (remaining 30% of sample cells).

Fig. 1.

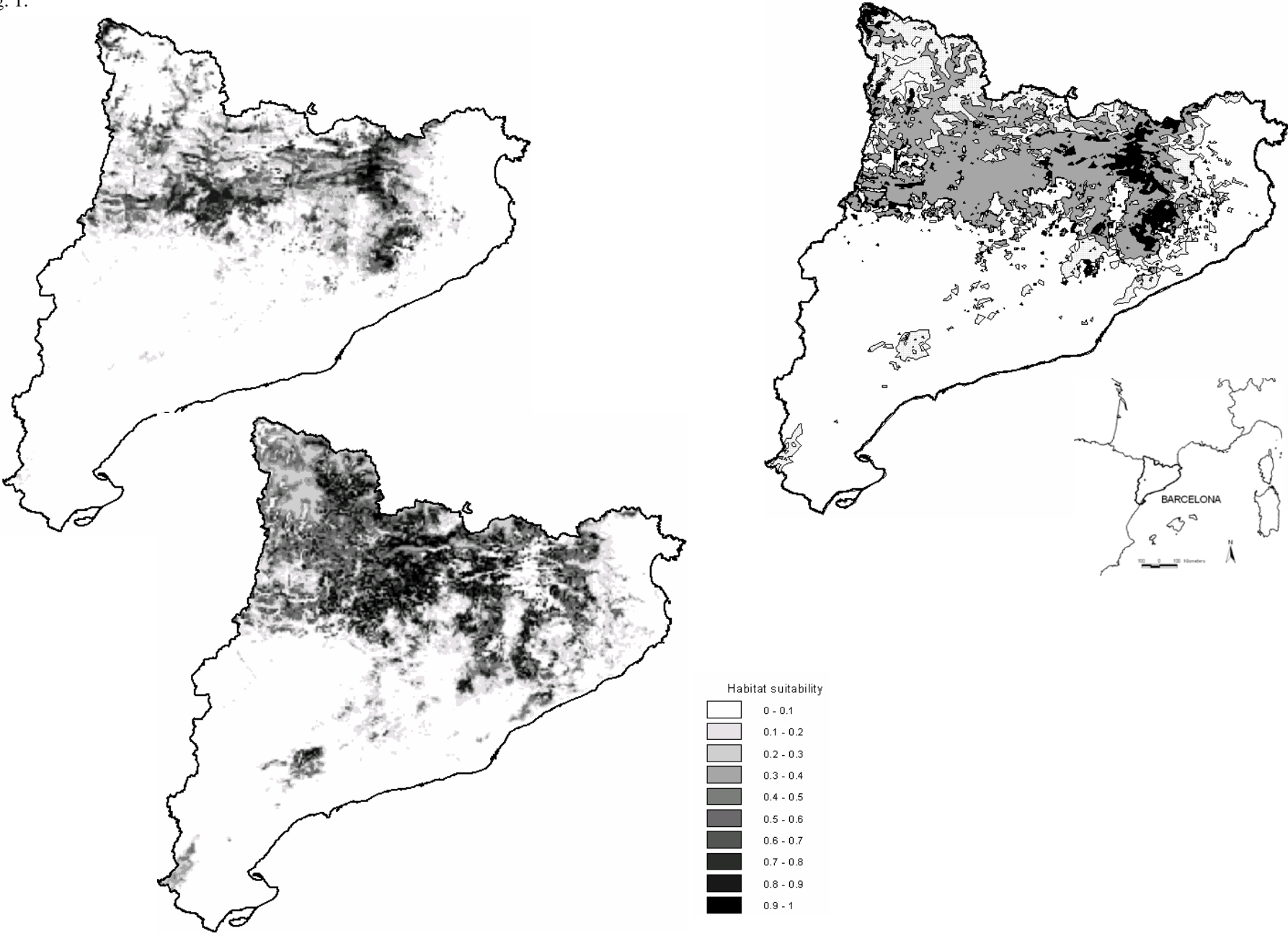


Fig. 2

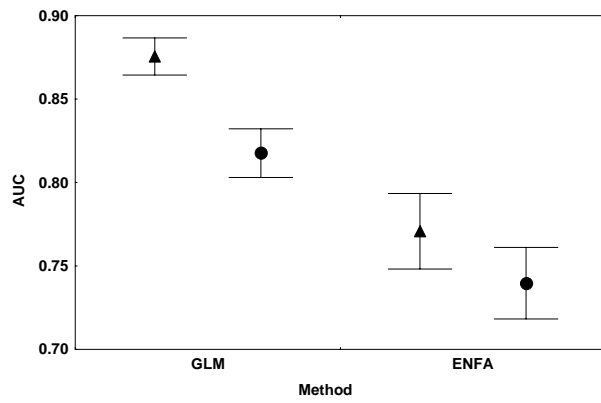


Fig. 3

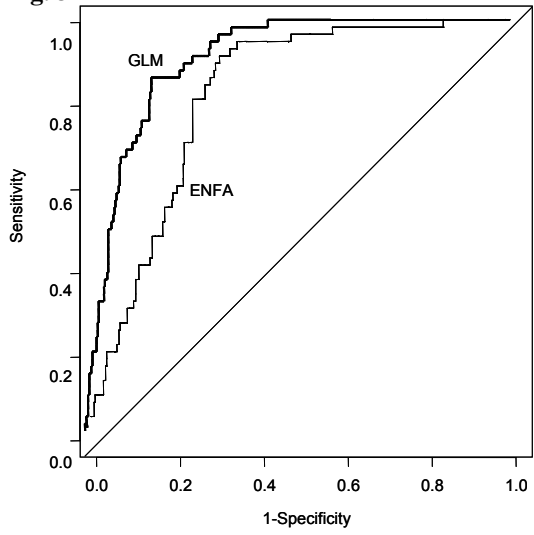
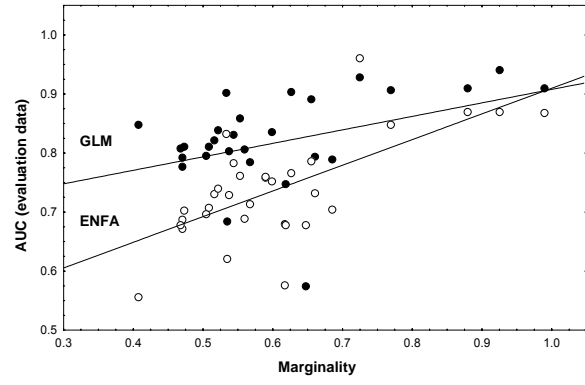
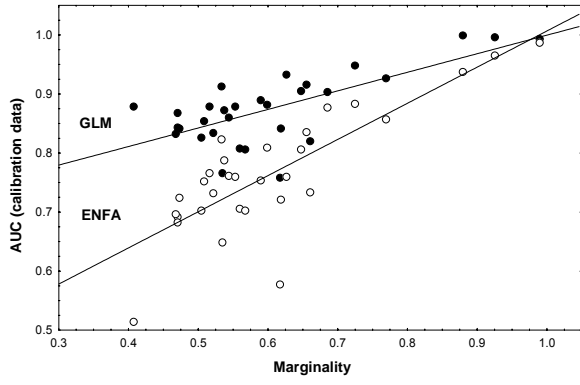
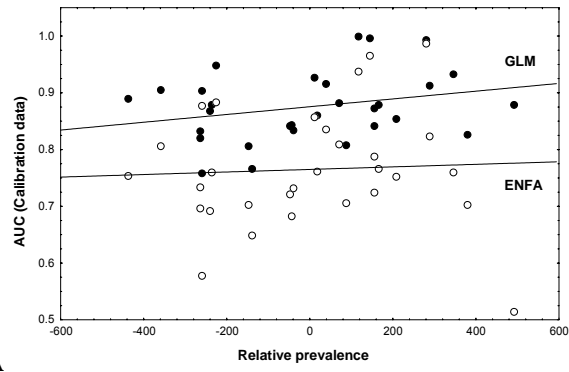


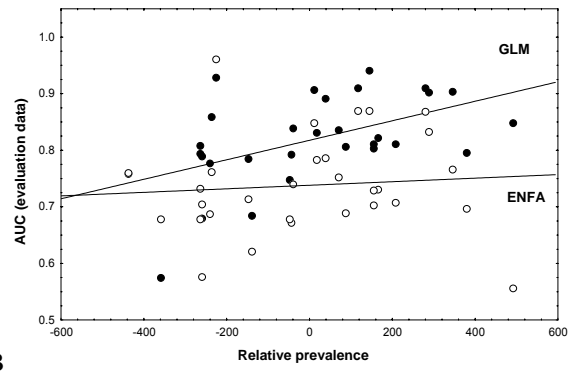
Fig. 4



A
Fig. 5



B



A

B

ANNEXE 1.4

Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain)

Thuiller W., Vaydera J., Pino J., Sabaté S., Lavorel S., Gracia C

Global Ecology & Biogeography (2003) **12**, 313-325



Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain)

WILFRIED THUILLER*†, JORDI VAYREDA‡, JOAN PINO‡, SANTI SABATE‡, SANDRA LAVOREL† and CARLES GRACIA‡

†Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de Recherche Scientifique, 1919 route de Mende, 34293 Montpellier cedex 5, France, E-mail: wilfried.thuiller@cefe.cnrs-mop.fr and ‡Centre de Recerca Ecològica i Aplicacions Forestals, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

ABSTRACT

Aim To explore the environmental correlates of tree species distributions in Catalonia according to the chorological status of the species.

Location The study area is the region of Catalonia, in north-eastern Spain.

Methods We used presence-absence data for 24 species, sampled in random plots distributed throughout forests of Catalonia. A climate model for the Catalonia region provided environmental variables. We used classification tree analysis to explore the environmental correlates of the realized niches of tree species. The predictive accuracy of the models was assessed using the ROC curve approach. Potential distribution maps of tree species were generated for the whole Catalonia region.

Results Models were ranked from low to high accuracy for the 24 species. Differences in accuracy among species were related to the chorological status of species. Zonal species, or

species at the core of their range (Mediterranean and Sub-Mediterranean species), were generally well predicted, while extrazonal species, or species at the edge of their range, were predicted only moderately well. Mediterranean species distributions showed good correlations with extreme temperatures and annual precipitation.

Main conclusions The above trends confirmed the difficulty of identifying the realized niche of species at the edges of their ranges. In contrast, Mediterranean and Sub-Mediterranean species, which were at the core of their range, were well-predicted, confirming the importance of extremes of temperature and annual precipitation as effective surrogates for variables having more direct physiological roles in limiting the ability of plants to survive and grow. Maps of potential tree distributions allowed us to define suitable habitats and to highlight areas where species have been planted outside their natural distribution.

Key words chorology, classification tree analysis, environmental gradients, forest trees, potential distribution maps, realized niche, Spain.

INTRODUCTION

The Mediterranean area is considered a hotspot of biodiversity (Myers *et al.*, 2000), mainly for its plant richness. Among plant species, trees are often considered as the basis for defining habitat representing many unsampled organisms. Conservation planning and projections of future distributions of plant and animal diversity therefore require efforts to determine the ecological requirements of forest tree species from their current distributions.

Tree species distributions are affected by biotic and abiotic

factors that define the realized niches of species. The realized niche is quantified from what determines the current spatial distribution of a species in a community or a region. Attempts to predict and explain species distributions should be based on the modelling of the realized niche (Austin & Smith, 1989). As the biotic factors limiting species can often not be measured over large scales, the realized niche is usually predicted using environmental variables taken as surrogates for physiologically relevant variables. Environmental variables might operate, however, at contrasting scales. Whereas large-scale climatic factors are usually responsible for the distribution of potential zonal communities, local topographic and geological factors determine the presence of extrazonal and azonal communities (Woodward, 1987; Ellenberg, 1988). In

* Corresponding author.

addition, distributions of all species are altered by human activity and natural disturbance. Large-scale climatic factors are commonly used to predict current plant distributions (Heikkinen & Birks, 1996; Burke, 2001; Frescino *et al.*, 2001) and to project future distributions, or species turnover in response to global change (Huntley *et al.*, 1995; Guisan *et al.*, 1998; Duckworth *et al.*, 2000; Sykes, 2001). However, at a finer resolution, climatic factors are likely to be overridden by local factors, which may be more relevant for explaining and predicting the current distributions of plants or animals (Huston, 1994). Finally, when working on limited geographical extents it is often difficult to capture the entire realized niche of species, because distributions span beyond the boundaries of study areas. In particular, if a species is at the edge of its range (i.e. extrazonal or azonal species), large-scale climatic factors may not explain its distribution, whilst more local-scale factors (topographical and geological) providing conditions prevailing outside of the study region (at the core of the species range) could be more relevant.

The present paper explores the relevance of biogeographical modelling of tree distributions in Catalonia (NE of Spain) in terms of their chorological status (Mediterranean, Sub-Mediterranean, Eurosiberian and Boreoalpine status) and in respect of the environmental variables potentially determining these distributions. Because of its geographical situation, at the borderline of the Eurosiberian and the Mediterranean phytogeographical regions, and containing Boreoalpine habitats in the Pyrenees range, Catalonia provides an interesting case study of the effects of biogeography on the accuracy of species modelling, with a great number of tree species with a variety of chorologies resulting in zonal, extrazonal and azonal situations. This analysis was made possible by the availability of data for forest tree diversity in Catalonia, collected as part of the Ecological Forest Inventory of Catalonia (IEFC), an extensive database of 10638 sampling points that covers all the forestry area in Catalonia at a density of 1 plot per km² (Gracia *et al.*, 2000a,b,c, 2001a,b).

This study addresses questions about the importance of environmental variables and forest type for plant survival and growth, and specifically examines which environmental variables are most suitable for modelling the realized niche of plant species in Catalonia. We apply classification tree analysis (CTA) in order to study the environmental correlates of the distributions of a set of important tree species. The comparison of the resulting models with current distribution maps is used to quantify the importance of environmental factors on the distribution of species according to their chorological status. Specifically we address the following questions:

- Are there differences of accuracy between models derived for species from different chorological classes?
- Are environmental requirements of species within a chorological class more similar than requirements of species in different classes?

MATERIALS AND METHODS

Study area

Catalonia is a region of the NE of Spain covering 32 098 km² (Fig. 1). The region exhibits a high climatic diversity resulting from its geomorphologic and topographical situation, receiving Mediterranean, Atlantic and even Saharan influences. The majority of the area has a Mediterranean climate, with a decrease in rainfall from north to south and from the coast to the central plains. A continentality gradient is also present from the coastal to inland areas, resulting in a gradual climatic variation from moist temperate coastal climates to dry continental inland climates. The north corresponds to the boundary of the Mediterranean and the Eurosiberian phytogeographical regions. The Pyrenees represent a sharp topographic-climatic gradient where Mediterranean or Eurosiberian-type biomes gradually change to subalpine and alpine types, corresponding to the Boreoalpine region.

Forest currently occupies 38% of Catalonia (Gracia *et al.*, in press), although its potential area covers up to 61%. Evergreen forests of *Quercus ilex* and *Q. suber* potentially constitute the late successional stage in Mediterranean lowland areas (Barbéro *et al.*, 1998). Deciduous oak (*Q. gr. humilis*) forests would substitute the evergreen forests in Sub-Mediterranean transition areas between the Mediterranean and the Eurosiberian regions. Eurosiberian to Boreoalpine forests are concentrated in the Pyrenees and their surrounding ranges, with deciduous forests of *Fagus sylvatica* and *Quercus petraea*, and Boreoalpine coniferous forests with *Abies alba* and *Pinus uncinata*.

The structure of Catalonian landscapes reflects a long and complex interaction between humans, forest and climate, with a historical background of forest regression since the beginning of the classic civilizations and a recent (over 100 year) reversion of this situation, with progressive crop abandonment, afforestation and forest recovery (Peix, 1999). Late successional species have been partially substituted by secondary forest species, mainly native conifers (*Pinus halepensis*, *P. pinaster*, *P. nigra*, *P. sylvestris*), but also alien ones (*P. pinea*, *P. radiata*), which represent the dominant fraction of Catalan forests. Large lowland forests are now dominated by *Pinus halepensis*, while *P. sylvestris* is the most abundant species in mountain areas. Natural disturbances also play a significant role in landscape structure, especially through frequent fires in the north-east of Catalonia.

Data set

The Ecological Forest Inventory of Catalonia (IEFC) is an extensive database of 10 638 circular sampling plots with a radius of 10 m. The field survey was carried out by the Centre for Ecological Research and Forestry Applications in Barcelona (CREAF). The IEFC includes a total of 95 tree

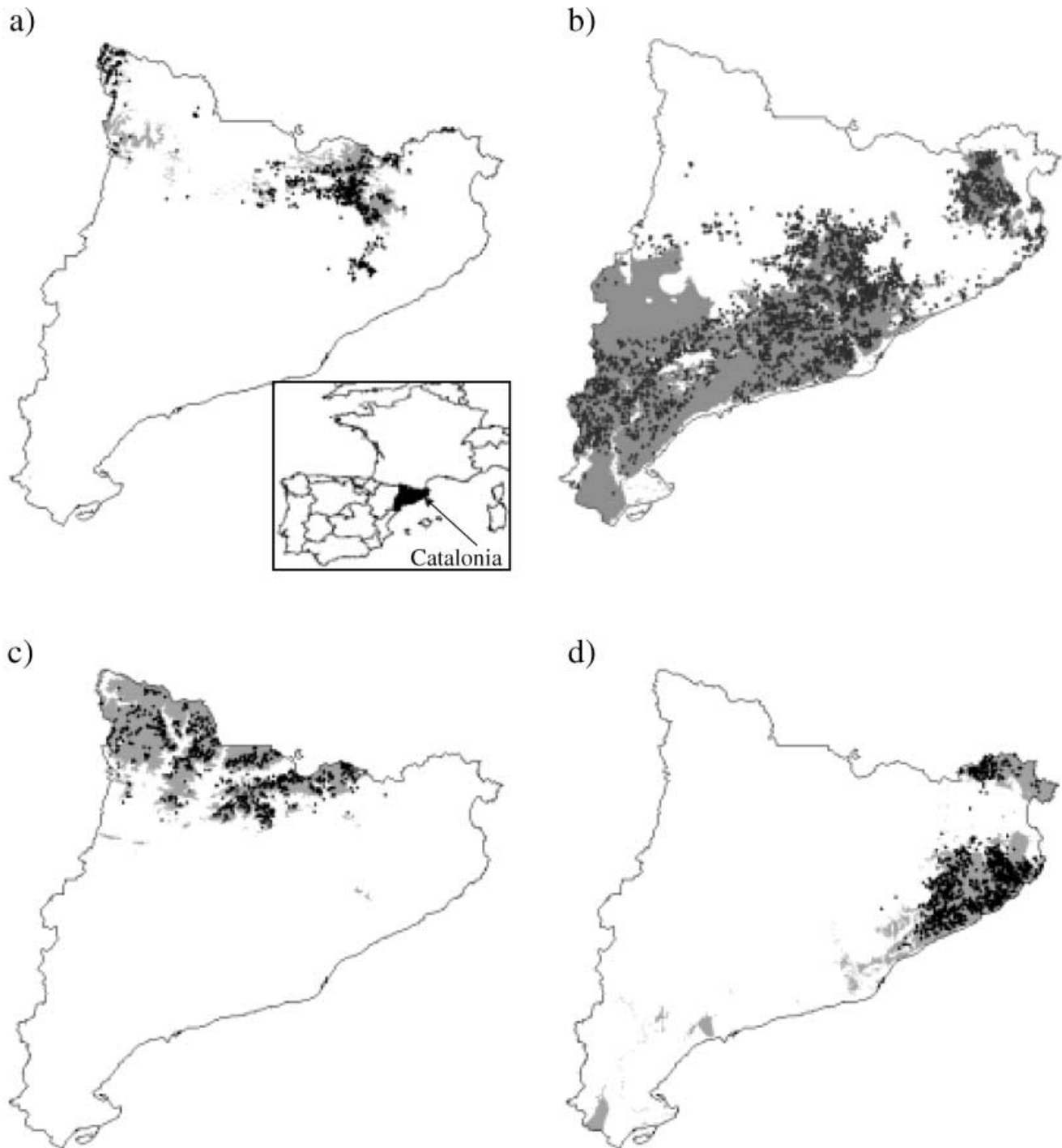


Fig. 1 Potential distribution map for the species very well modelled by CTA. (a) *Fagus sylvatica* (b) *Pinus halepensis* (c) *Pinus uncinata* and (d) *Quercus suber*. Points represent the sampling plots where the species was present and shaded areas are the areas modelled as suitable for the species.

Table 1 List of the selected environmental variables

Environmental variables	Abbreviation
Annual Precipitation	MAP
Summer Precipitation	SumPrec
Winter Precipitation	WintPrec
Autumn Precipitation	AutPrec
Spring Precipitation	SprPrec
Mean Annual Temperature	MAT
Mean Temperature of the Coldest Month (January)	MTC
Mean Temperature of the Hottest Month (July)	MTW
Mean Annual Maximum Temperature	MAMxT
Mean Maximum Temperature of Hottest Month (July)	MMTW
Mean Annual Minimum Temperature	MAMiT
Mean Minimum Temperature of the Coldest Month (January)	MMTC
Mean Annual Radiation	MAR
Slope	Slope
Geology (factor)	Geol
category 1: Siliceous rocks	
category 2: Carbonated rocks	
category 3: Limestone	
category 4: Calcareous sediments	
category 5: Siliceous sediments	

species (of which 5 are present in more than 10% of the plots), including the most abundant *Pinus* (7 spp.) and *Quercus* (12 spp.) species.

For this study, we selected a set of 24 forest tree species representative of the compositional diversity of Catalonia. This set includes common, intermediate and less common species, from 12 genera: *Quercus* (4 spp.), *Pinus* (6 spp.), *Acer* (2 spp.), *Juniperus* (2 spp.), *Populus* (2 spp.), *Abies*, *Castanea*, *Corylus*, *Fagus*, *Salix*, *Sorbus* and *Ulmus* (1 sp.). These species are also a representative sample of the existing variety of phytogeographical patterns, including Boreoalpine (2 spp.), Eurosiberian (10 spp.), Sub-Mediterranean (6 spp.) and Mediterranean (6 spp.) species. These different phytogeographical patterns correspond with species at the edge of their distribution ranges (Boreoalpine and Eurosiberian) compared to species at the core of their ranges (Mediterranean and sub-Mediterranean).

Environmental variables

Environmental variables were selected according to their relevance to tree survival and growth (Table 1). Although the main environmental data required were climate data, given the gradients of altitude and topography in Catalonia, we also included slope and geology. Climatic variables were taken as surrogates for variables having more direct physiological roles in limiting the ability of plants to survive and grow, such as the number of growing degree days or actual evapotranspiration (Bartlein *et al.*, 1986; Prentice *et al.*, 1991; Huntley *et al.*, 1995). Geology is an indirect variable

and was considered as a surrogate for soil type, with direct impacts on nutrient and water availability for plant growth (Austin & Smith, 1989).

Climatic variables (temperature and precipitation) were generated by a climate model (Ninyerola *et al.*, 2000). This was used to generate layers of monthly precipitation and monthly mean, minimum and maximum temperatures for the entire study area. The climate model used a network of weather stations in Catalonia (257 stations for precipitation and 160 stations for temperatures), and was based on a multiple regression analysis between the meteorological variables and a set of geographical variables (latitude, solar radiation, altitude, continentality and cloudiness factors). Layers of rainfall and temperature were then derived from the original climate model (Ninyerola *et al.*, 2000). Topographic layers (elevation, slope and solar radiation) were provided by a DEM (180-m resolution) generated by the Cartographic Institute of Catalonia (ICC).

The data were used in the raster format of ARCVIEW (ESRI Inc., Redlands, California). We combined raster and sampling locations for each species data to determine the values of each environmental variable in each plot.

Statistical analysis

Distribution data sets

Models were calibrated using a random sample of the data (70%: calibration data) and evaluated with the remaining 30% (evaluation data).

Model

We used classification tree analysis (*tree* in Splus software: Anon, 1999) to study and analyse the relationships between environmental factors and species distributions (Breiman *et al.*, 1984). The tree is built by repeatedly splitting the calibration data, according to a simple rule based on a single explanatory variable. At each split, the data are partitioned into two exclusive groups, each of which is as homogeneous as possible. The heterogeneity of a node is defined with a deviance notion that can be interpreted as the deviance of a Gaussian model (regression tree) or multinomial model (classification tree) (Breiman *et al.*, 1984). To control the length of the tree, we used the 'prune' function of S-plus (Anon, 1999). The program builds a nested sequence of subtrees of the main tree by recursively snipping off the less important splits in terms of explained deviance.

CTAs have only recently received increased interest for biogeographical studies (Franklin, 1998; Iverson & Prasad, 1998; Vayssières *et al.*, 2000; Rouget *et al.*, 2001) as compared to other modelling techniques like generalized linear and generalized additive models (GLM and GAM, respectively). Several studies have shown that CTA offer a good compromise, being as accurate as GLM or GAM (Franklin, 1998; Vayssières *et al.*, 2000), and offering the following useful properties:

Facility of implementation: CTAs accept mixes of categorical and continuous data (Breiman *et al.*, 1984; Iverson & Prasad, 1998), as also possible with GLM, but not with GAM.

Tree representation provides a hierarchical view of the relationships between species and environmental variables and makes it possible to identify which variable is the major correlate of the presence of a species (Iverson *et al.*, 1999; Vayssières *et al.*, 2000).

CTAs are one of the few methods that can model interactive effects of two or more variables, and can represent them in an easily readable fashion (Iverson & Prasad, 1998; Rouget *et al.*, 2001). Interactions factors can also be added to GLM, but with an *a priori* choice of interactions or using stepwise procedures. Conversely, GAM is purely additive and cannot deal explicitly with interaction terms.

Assessing prediction accuracy

We used the Receiver Operating Characteristic curve (ROC curve) to evaluate the accuracy of our models (Hanley & McNeil, 1982, 1983). This graphical method makes it possible to represent the relationship between the percentage of presences correctly predicted (sensitivity) and 1 minus the percentage of the absences correctly predicted (specificity). The area comprised between a 45° line and the curve measures the discrimination, that is, the ability of the model to classify correctly a species as present or absent in a given plot

(area under the curve: AUC). The confidence intervals of AUC at 95% were derived for each species (Pearce & Ferrier, 2000; Vayssières *et al.*, 2000). A rough guide for classifying the accuracy is: 0.50–0.60 = fail; 0.60–0.70 = poor; 0.70–0.80 = fair; 0.80–0.90 = good; 0.90–1 = excellent (Swets, 1988).

Potential distributions

Given the response surface fitted to the observed patterns of occurrence in the sampled plots and the environmental variables for the 24 selected species, it is possible to map the potential distributions of these species for the entire Catalonia region. To plot potential distributions of species, probability values of presence from CTAs were converted into binary form using a threshold maximizing the sensitivity and specificity of the predictions (Pearce & Ferrier, 2000) for each species.

RESULTS

Accuracy of classification tree analysis

We considered four classes of model accuracy (Table 2). The low accuracy (AUC < 0.7) class included *Juniperus communis* and *Ulmus minor* (AUC = 0.646 and AUC = 0.645, respectively). *Ulmus minor* had the lowest accuracy of all species. Six species populated the fair accuracy class (0.7 < AUC < 0.8), including species with many occurrences (*Quercus humilis*, *Q. ilex*) and some scarce species (*Salix caprea*, *Ilex aquifolium*). Good accuracy (0.8 < AUC < 0.9) was assigned to 10 species belonging to 7 genera, including some highly abundant (e.g. *Pinus sylvestris*, 3217 occurrences) and other scarcer species (e.g. *Acer opalus*, 240 occurrences). Within this class two subgroups were considered according to levels of accuracy (according to AUC and confidence intervals). A first subgroup (*Salix caprea*, *Pinus pinaster*, *Juniperus oxycedrus*, *Quercus petraea*, *Sorbus aria* and *Acer opalus*) had AUC ranging from 0.808 to 0.835, and a second subgroup (*Pinus sylvestris*, *Populus tremula*, *Pinus nigra* and *Pinus pinea*) had AUC ranging from 0.875 and 0.891. The high accuracy class (AUC > 0.9) included six species (*Castanea sativa*, *Abies alba*, *Pinus halepensis*, *Fagus sylvatica*, *Quercus suber* and *P. uncinata*). *P. uncinata* was the best-predicted species, with a narrow confidence interval (0.977 < AUC < 0.982) and an extremely high percentage of presences (90.6%) and absences (94.7%) correctly predicted. Accuracy of CTA predictions increased from Eurosiberian species, with only 50% of species being scored in high or good accuracy classes, to Mediterranean ones, which showed 83% of species in these classes. Sub-Mediterranean trees exhibited an intermediate situation, with 67% of species falling into high or good accuracy classes.

Table 2 Prediction accuracy for the 24 species

Species	Chorology	<i>n</i>	Calibration data		Evaluation data				
			AUC		Pres	Abs	AUC	Lo. 0.95	Up. 0.95
High accuracy									
<i>Pinus uncinata</i>	B	787	0.976		90.6	94.7	0.979	0.977	0.982
<i>Quercus suber</i>	M	1034	0.981		92.0	93.6	0.956	0.947	0.966
<i>Fagus sylvatica</i>	E	567	0.958		85.8	85.4	0.941	0.933	0.949
<i>Pinus halepensis</i>	M	2712	0.950		87.1	86.8	0.941	0.936	0.945
<i>Abies alba</i>	B	229	0.979		93.0	89.6	0.937	0.915	0.959
<i>Castanea sativa</i>	SM	246	0.957		90.0	82.5	0.932	0.918	0.946
Good accuracy									
<i>Pinus pinea</i>	M	997	0.892		88.8	74.8	0.891	0.883	0.899
<i>Pinus nigra</i>	SM	1971	0.910		89.6	75.5	0.890	0.888	0.897
<i>Populus tremula</i>	E	151	0.899		75.7	84.9	0.886	0.864	0.904
<i>Pinus sylvestris</i>	SM	3217	0.898		75.8	78.7	0.875	0.868	0.882
<i>Acer opalus</i>	SM	240	0.907		78.7	77.8	0.835	0.814	0.857
<i>Sorbus aria</i>	E	140	0.888		77.1	80.0	0.832	0.796	0.869
<i>Quercus petraea</i>	E	262	0.878		81.4	71.2	0.817	0.787	0.848
<i>Juniperus oxycedrus</i>	M	146	0.815		65.8	86.3	0.813	0.775	0.8
<i>Pinus pinaster</i>	M	302	0.848		78.5	81.2	0.81	0.783	0.836
<i>Salix caprea</i>	E	91	0.914		79.3	81.0	0.808	0.753	0.863
Fair accuracy									
<i>Quercus humilis</i>	SM	1847	0.825		79.0	68.7	0.797	0.786	0.808
<i>Populus nigra</i>	E	102	0.882		59.4	86.5	0.796	0.751	0.841
<i>Acer monspessulanum</i>	SM	157	0.885		64.9	72.1	0.785	0.750	0.820
<i>Quercus ilex</i>	M	4070	0.786		74.5	68.7	0.779	0.771	0.788
<i>Ilex aquifolium</i>	E	102	0.856		76.2	71.6	0.765	0.730	0.800
<i>Corylus avellana</i>	E	182	0.919		73.7	63.5	0.752	0.709	0.796
Low accuracy									
<i>Juniperus communis</i>	E	212	0.819		68.9	59.1	0.646	0.610	0.682
<i>Ulmus minor</i>	E	115	0.942		63.9	65.2	0.645	0.592	0.703

Abbreviations for chorology: B, Boreoalpine; E, Eurosiberian; M, Mediterranean; SM, Sub-Mediterranean. *n* = number of occurrences in the sample sites. Pres: mean percentage of presences correctly predicted. Abs: mean percentage of absences correctly predicted. AUC, Area Under the Curve of the Relative Operating Characteristic Curve (ROC curve). Lo. 95 and Up. 95 represent the lower and upper limits of the confidence interval of AUC.

Potential distribution area of the species

Species with highly accurate predictions exhibited potential distributions very close to the observed ones, with almost all observed points enclosed in the potential area (Figs 1, 2 and 3). Well-modelled (Fig. 1) species included contrasting types of distributions, ranging from widespread species such as *Pinus halepensis* to localized species such as *Pinus uncinata* and *Abies alba*, which grow exclusively in the Pyrenees, and *Quercus suber*, which is limited by siliceous areas in the north-east. Dominant species belonging to good and fair accuracy classes (*Pinus nigra*, *P. sylvestris*, *Quercus ilex*, *Q. humilis*) had similar potential and observed distributions (Fig. 2), although there were many outliers, mainly in peripheral areas. In the case of *Quercus ilex*, the model did not predict accurately the inland distribution of

the species. There were still greater differences between potential and observed distributions for several minor species of these two accuracy classes (Fig. 3), probably because of the limited number of sampling points that did not permit an accurate modelling in large parts of their potential area of distribution.

Environmental correlates for the distribution of well-modelled species

Classification trees for several well-modelled species (*Fagus sylvatica*, *Pinus uncinata*, *Pinus halepensis* and *Quercus suber*) are shown in Fig. 4. Tree-based models can identify and express in a relatively simple form nonlinear and non-additive relationships with quantitative thresholds from presence to absence. For instance, altitude above 1335 m was

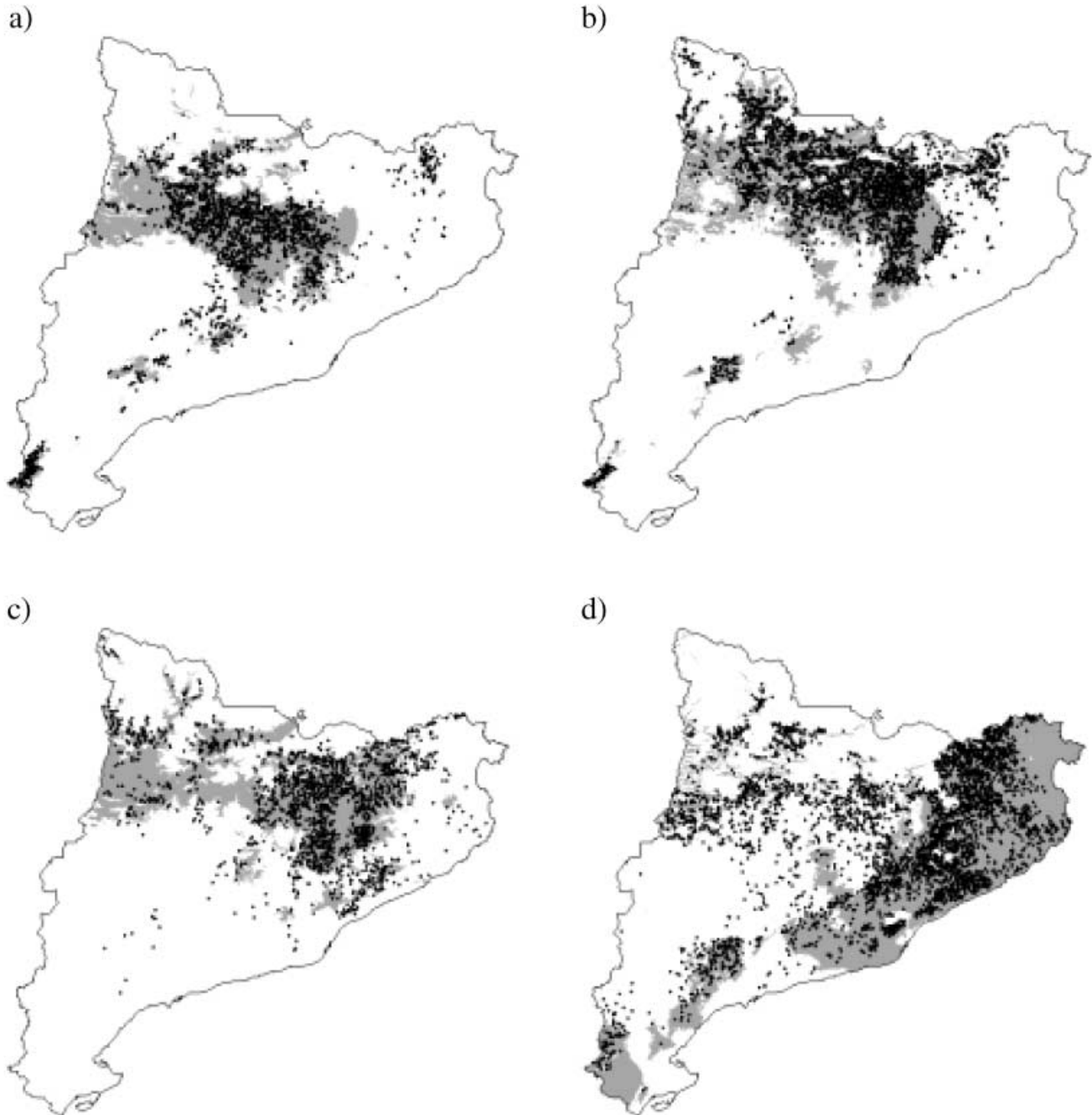


Fig. 2 Potential map of species well modelled by CTA. (a) *Pinus nigra* (b) *Pinus sylvestris* (c) *Quercus humilis* and (d) *Quercus ilex*. Points represent the sampling plots where the species was present, and shaded areas are the areas modelled as suitable for the species.

identified as the main correlate for the presence of *P. uncinata*, but temperature and rainfall appeared as important factors in the lower part of this range below 1720 m. Mean temperature of the coldest month was identified as the main limiting factor for *Quercus suber*, while siliceous soil came second.

The model predicted the presence of the species on calcareous soil only in high precipitation areas. The presence of *Pinus halepensis* was mainly correlated with annual precipitation, although this apparent constraint might be compensated by the interaction with many other variables, including other

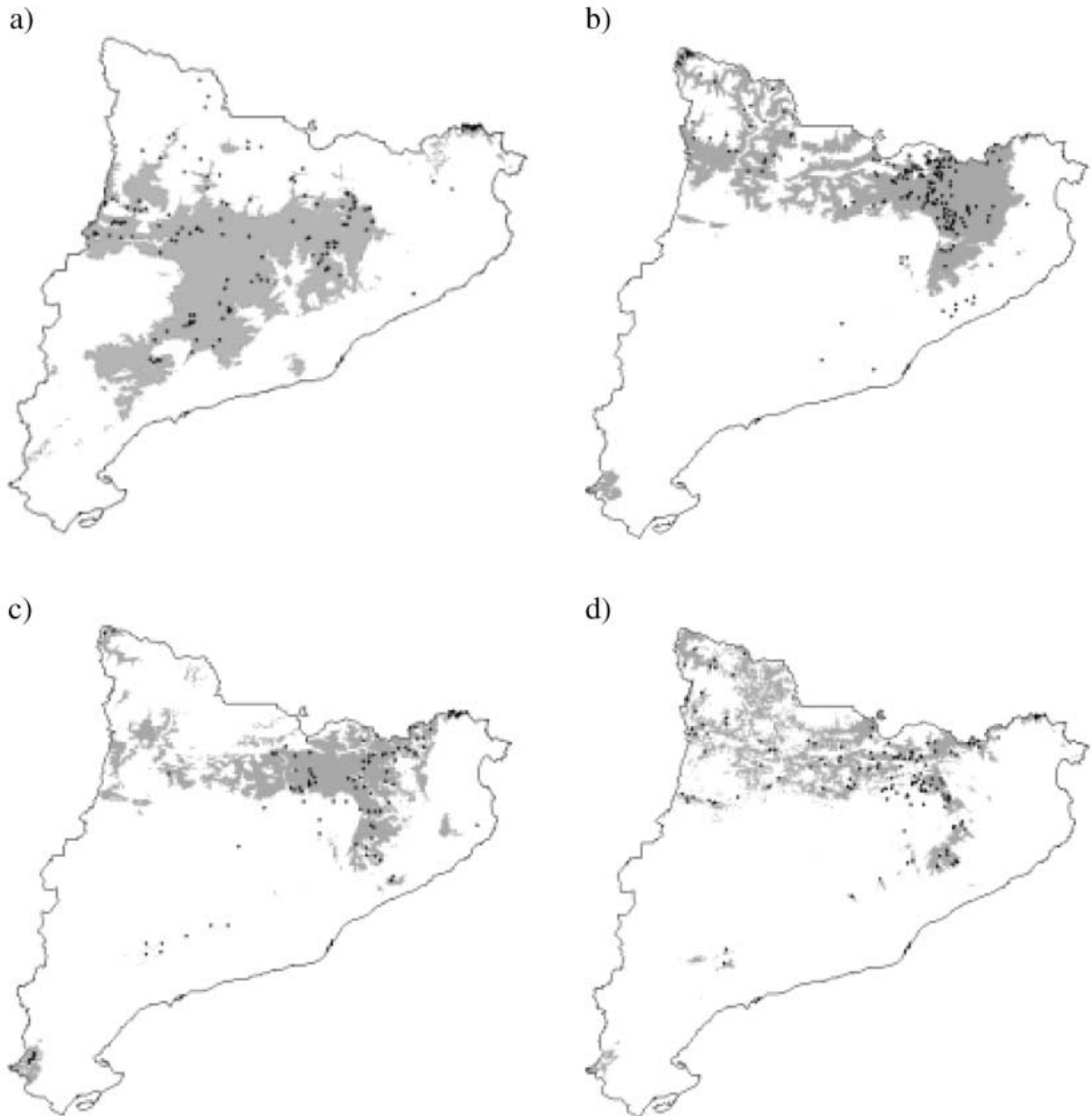


Fig. 3 Potential map of the rare species fairly modelled by CTA. (a) *Acer monspessulanum* (b) *Corylus avellana* (c) *Ilex aquifolium* and (d) *Sorbus aria*. Points represent the sampling plots where the species was present, and shaded areas are the areas modelled as suitable for the species.

components of precipitation, temperature and geology. *Fagus sylvatica* appeared mainly restricted to areas with > 950 mm of annual rainfall, within which, in areas with less than 1050 mm, the distribution was related to winter and summer precipitation.

Main correlates of tree species distributions

The models most often selected six variables as important variables (Table 3): three precipitation variables (annual, summer and winter), two temperature variables (temperature

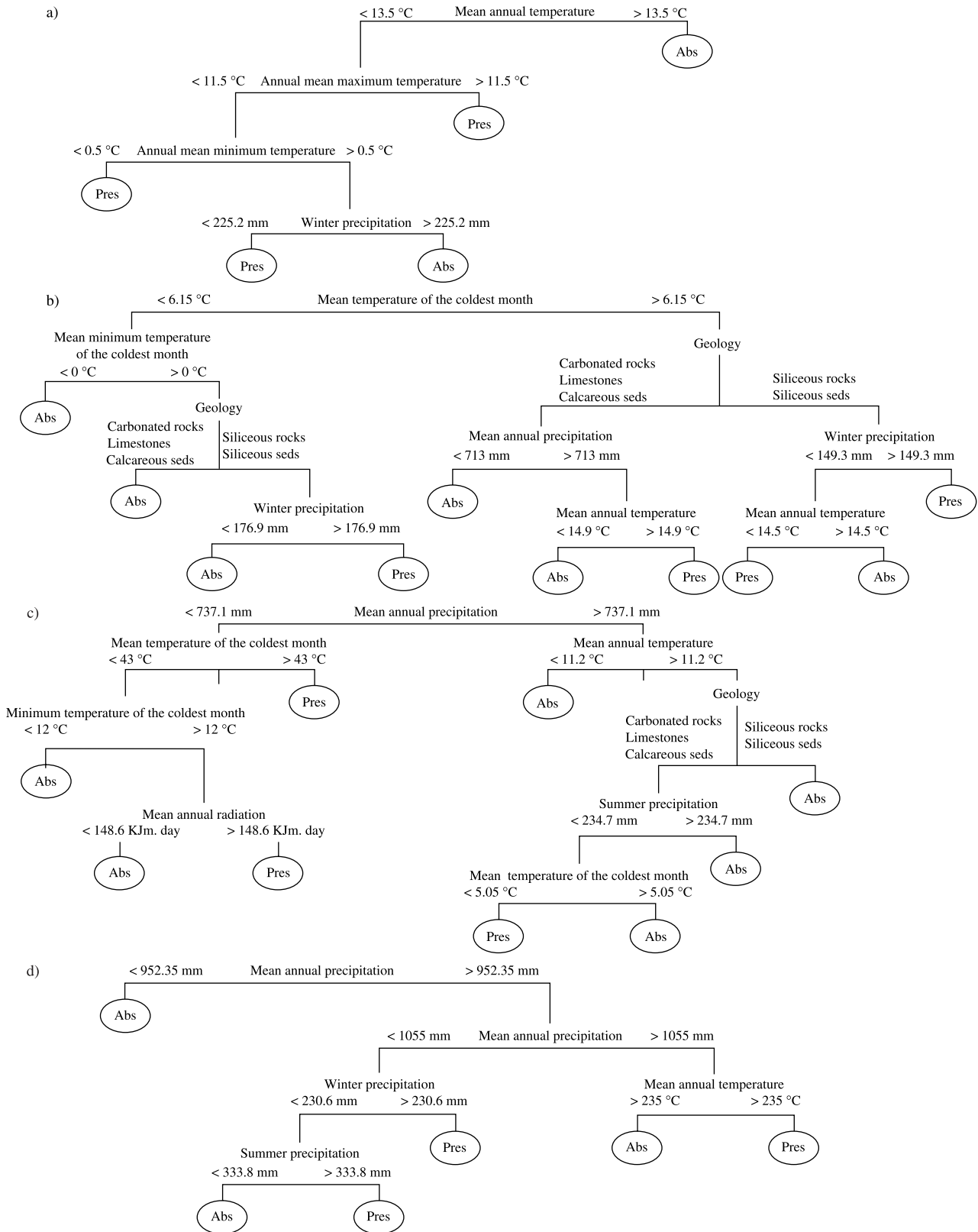


Fig. 4 Main predictors of: (a) *Pinus uncinata* (b) *Quercus suber* (c) *Pinus halepensis* and (d) *Fagus sylvatica* distribution as modelled by CTA. The combination of variables that favour the occurrence of a species are represented by Pres. The dendrogram uses the change in deviance to guide the vertical positioning of each pair of nodes.

Table 3 List of the first four environmental variables (in terms of explained deviance) selected by the classification tree analysis for each species. The last row presents the relative importance of each variable for the 24 species. For each species, an index has been attributed to the variables following the order of selection by the model. The value of 4 denotes the first variable, 3 the second, 2 the third and 1 the fourth. Phytogeographical classification of the species is according to Bolòs *et al.* (1990). See Table 1 for variable abbreviations

	MAP	Win Prec	Spr Prec	Sum Prec	Aut Prec	MAT	MA MiT	MTC	MMTC	MAMxT	MTW	MM TW	MAR	Slope	Geology
Boreoalpine															
<i>Abies alba</i>		2		3							1		4		
<i>Pinus uncinata</i>		3					2			1					
Eurosiberian															
<i>Corylus avellana</i>	1		2			3									4
<i>Fagus sylvatica</i>	1	3		4						2					
<i>Ilex aquifolium</i>	1							3			4		2		
<i>Juniperus communis</i>		3		1	4		2								
<i>Populus nigra</i>		4										3		1	2
<i>Populus tremula</i>	1		2							3			4		
<i>Quercus petraea</i>			1	3						2		4			
<i>Salix caprea</i>				2		1			4						3
<i>Sorbus aria</i>					3					4	1			2	
<i>Ulmus minor</i>				2				1		3			4		
Submediterranean															
<i>Acer monspessulanum</i>	4				1	2							3		
<i>Acer opalus</i>			3						4		1				2
<i>Castanea sativa</i>	3								2				4		1
<i>Pinus nigra</i>		1	3	4			2								
<i>Pinus sylvestris</i>	3	4		1		2									
<i>Quercus humilis</i>				1	4				2	3					
Mediterranean															
<i>Juniperus oxycedrus</i>					1			2	4			3			
<i>Pinus halepensis</i>	1							2	3				4		
<i>Pinus pinaster</i>		2			4		3	1							
<i>Pinus pinea</i>	3			4			2	1							
<i>Quercus ilex</i>		2		3						1					4
<i>Quercus suber</i>		4						1	2						3
First position	5	1	1	3	2	1	0	4	0	2	3	0	0	1	1
Second position	0	3	2	2	0	2	4	2	3	2	1	0	1	1	2
Third position	3	3	2	3	1	1	1	1	1	3	0	2	1	0	2
Fourth position	1	3	0	2	2	0	0	0	3	4	4	1	4	0	2
Importance of variables	27	22	14	26	12	12	14	24	14	24	19	5	9	7	16

of the coldest month and maximum annual temperature), and geology. Among these six main variables, annual precipitation and mean temperature of the coldest month were most frequently selected as primary factors (in first or second position), while geology, winter and summer precipitation were selected at a variable rank. These six variables had varying weights across species, even within a given genus. For example, the models selected mean winter precipitation as a main determinant for the distribution of all *Pinus* species, except for *P. halepensis*, which appeared to be mainly affected by annual precipitation. Mean temperature of the coldest month was a main environmental constraint for *P. pinea*, *P. pinaster* and *P. halepensis*, but not for the remaining species of the genus. Table 3 reveals patterns within chorological classes

rather than taxonomic classes. Mediterranean species are distributed mainly in relation to temperature and secondly are related to precipitation. In contrast, Sub-Mediterranean and Eurosiberian species seem to be influenced mainly by precipitation and secondly by temperature. It is more difficult to find a general pattern for the two Boreoalpine species, both of which are at the edge of their range.

DISCUSSION

Classification trees

As already depicted by Franklin (1998), Iverson & Prasad (1998) and Vayssières *et al.* (2000), CTA provides an interesting

tool for analysing the main correlates of species distribution. CTA displays a hierarchical easily understood view of species-environment relationships. Several studies have shown their relatively good predictive performance (Harvey, 1996; Vayssières *et al.*, 2000) and most of them underline the power of CTA to find interactions and hierarchical relations among environmental variables (Hastie & Tibshirani, 1990; Austin, 2002). However, generalized linear and additive models may be more powerful tools, providing more meaningful relationships between species and explanatory variables (e.g. bell-shaped curves) (Austin, 2002). As already shown by several studies, there is no universal method and the choice of a modelling method should be made according to the aim of the study. In our case, CTA seemed to be the more suitable tool. Alternatively, dynamic vegetation models (Prentice & Leemans, 1990; Sykes & Prentice, 1996) incorporate explicit representations of key ecological processes (tree growth, establishment, competition, succession) and have been widely used to simulate the potential response of vegetation to past and future climate change (Prentice *et al.*, 1991; Sykes & Prentice, 1995; Korzukhin *et al.*, 1996; Starfield & Chapin, 1996). However, since such models require extensive knowledge of the species involved, as well as of the historical factors of the studied area, they can only be developed for well-studied species and habitats (Peng, 2000). In our case, perfect knowledge of species ecology and historical factors over the entire Catalonia region were missing, precluding the use of such approaches.

Correlates of tree species distribution

Classification Tree Analyses succeeded as a method of identifying the main correlates of actual species distributions and as a method for predicting the potential distributions of 67% (high or good accuracy) of the selected species. Model accuracy increased from Eurosiberian species (50% of species scored as high or good accuracy classes) through Sub-Mediterranean (67%) to Mediterranean (83%). Trends in accuracy mirrored the decreasing extrazonality or azonality from Eurosiberian (species at the edge of range) to Mediterranean species (species at the core of range) (Ellenberg, 1988). Catalonia is located in the borderline of the Eurosiberian and the Mediterranean regions and the general climate is Mediterranean. We confirmed the importance of annual precipitation and extreme temperatures in explaining regional patterns of distribution for Mediterranean species, in accordance with other studies at similar scales (Barbéro *et al.*, 1998; Rouget *et al.*, 2001). For example, *Pinus halepensis*, which belongs to the Klaus (1989) group of Mediterranean shore and island pines, is expected to occur in semihumid to semiarid habitat (Barbéro *et al.*, 1998). Our model corroborated such expectations, showing that the distribution of *P. halepensis* is related to intermediate annual precipitation values and is restricted

to mild winter temperatures (Fig. 4). This species is well adapted to Mediterranean conditions and the study area is situated at the core of its distribution, making for accurate predictions of its potential distribution in Catalonia. Conversely, Eurosiberian and Borealpine species at the edge of their range are restricted to specific local climatic and topographic situations, where landscape factors (aspect, convexity, water table height) can be much more important than the average climate. This is particularly true for riparian species such as *Populus nigra* and *Ulmus minor*, which therefore showed fair to low model accuracies. The two Borealpine species (*Abies alba* and *Pinus uncinata*) are restricted to the Pyrenees and have specific local requirements (climatic and topographic), hence our relatively good predictions. The relative failure for species at the edge of their distribution range could be improved by working at a finer scale and with more topographic and geomorphologic variables (Reed *et al.*, 1993; Bridge & Johnson, 2000). However, models can also show high accuracy for some species at the edge of their range, like *Fagus sylvatica* (AUC = 0.94). This species is continental and occurs principally in central Europe, Catalonia being its southern limit. Huntley *et al.* (1989) and Prentice & Helmisaari (1991) showed that *F. sylvatica* is found mainly in sites with annual precipitation close to 1200 mm and is absent from places where the mean temperature of the coldest month is less than -3°C . Our study suggested the same restrictions due to cold limits and annual precipitation, with values close to those reported by Prentice & Helmisaari (1991).

Finally, there is no general pattern in terms of accuracy for azonal species. Further developments on the chorology and modelling of numerous azonal species are needed to find a straightforward explanation of such patterns.

Land use and biogeographical issues

Logically, the best-modelled species were those growing in a large proportion of their potential habitats. This is the case for *Pinus halepensis*, which, thanks to its high colonizing ability, is only absent from its potential distribution area in the arid inland zone, devoted to agriculture, and in the extreme south-east, as a result of frequent wildfires. Other well-modelled species are dominant trees in forests growing in relatively extreme habitats where human impact has been traditionally low. This is the case for subalpine forests of *Pinus uncinata* and *Abies alba*, and to a lesser degree for mountain forests of *Fagus sylvatica*, which are all located in cold and moist north-facing areas unsuitable for agriculture. The same can also apply to *Quercus suber* and *Castanea sativa*, growing in siliceous, poor soils, which are also inadequate for cropping.

Model accuracy was lower for the remaining species. Among them, forestry pine species (*Pinus nigra*, *P. pinea*, *P. sylvestris*, and *P. pinaster*) were modelled with good accuracy, indicating that they have been planted or have spread mainly

within their potential areas. Although we would expect that their current distributions would be strongly influenced by land use history and current management, in the absence of data for these variables bioclimatic models seemed sufficient. These commercial species have been planted at a range of sites known to be climatically suitable. Consequently, climate may provide a reasonable surrogate for land management effects. However, models rightly detected as false negatives occurrences of *P. sylvestris* and *P. nigra* in north-eastern Catalonia, which has a maritime climate, and where these species are not native but planted (Fig. 2). False negatives typically reflect the inability of static equilibrium models such as CTA to deal with historical factors and dynamic behaviour involved when species are planted outside their natural environmental range (Guisan & Zimmermann, 2000).

Minor Eurosiberian species were in the good to low accuracy categories, indicating a variable dependence on average environmental conditions. Some of these species (*Populus tremula*, *Sorbus aria* and *Ilex aquifolium*) are found in large forests of southern Catalonia (Pino *et al.*, 2001), at the border of their range, where only small and localized populations persist. In these areas they behave as forest interior species (Forman, 1998) where they are protected from interactions with adjacent human habitats. Alternatively the fair or poor accuracy of models for such azonal species could be a result of the sampling strategy of the IEF. Sampling favours abundant species and scarce species are likely to have been under-sampled. This explanation is also plausible for widespread but scarce species like *Acer monspessulanum* or *Sorbus aria*. *Q. ilex* and *Q. humilis* make up a third group of species with models of fair accuracy. *Q. ilex* forests would be the dominant vegetation in non siliceous lowland areas in Catalonia and southern France, and *Q. humilis* forests would dominate the Sub-Mediterranean areas bordering the Eurosiberian region (Bolòs *et al.*, 1990; Barbéro *et al.*, 1998). However, in the majority of their potential area they have been substituted by pine forests, scrublands, grasslands, croplands and urban areas and therefore occupy a small proportion of their potential distribution. This pattern is the likely cause for the low accuracy of models for these species. In addition the model did not accurately predict the inland distribution of *Q. ilex*, where the typical subspecies *Q. ilex* ssp. *ilex* is substituted by another (*Q. ilex* ssp. *ballota*) with contrasting ecological requirements.

ACKNOWLEDGMENTS

This research was funded by the European Union's ATEAM (EVK2-CT-2000-00075) project. Species and environmental data were kindly made available to us by the Centre for Ecological Research and Forestry Applications in Barcelona. We are indebted to Lluís Brotons, Miguel Araújo, John Thompson, Martin Sykes and two anonymous reviewers for comments

on successive versions of the manuscript. This paper contributes to the Montpellier — Barcelona European Laboratory (LEA, CNRS) 'Mediterranean Ecosystems in a Changing World'.

REFERENCES

- Anon (1999) *S-PLUS 2000 User's Guide*. Data Analysis Products Division, MathSoft, Seattle.
- Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.
- Austin, M.P. & Smith, T.M. (1989) A new model for the continuum concept. *Vegetatio*, **83**, 35–47.
- Barbéro, M., Loisel, R., Quézel, P., Richardson, D.M. & Romane, F. (1998) Pines of the Mediterranean Basin. *Ecology and biogeography of Pinus* (ed. by D.M. Richardson), pp. 450–473. Cambridge University Press, Cambridge.
- Bartlein, P.J., Prentice, I.C. & Webb, T. (1986) Climatic response surfaces from pollen data for some eastern North American taxa. *Journal of Biogeography*, **13**, 35–57.
- Bolòs, O.D., Vigo, J., Masalles, R.M. & Ninot, J.M. (1990) *Flora manual dels països. Catalans* Pòrtic, Barcelona.
- Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984) *Classification and regression trees*. Chapman & Hall, New York.
- Bridge, S.R.J. & Johnson, E.A. (2000) Geomorphic principles of terrain organization and vegetation gradients. *Journal of Vegetation Science*, **11**, 57–70.
- Burke, A. (2001) Classification and ordination of plant communities of the Naukluft Mountains, Namibia. *Journal of Vegetation Science*, **12**, 53–60.
- Duckworth, J.C., Bunce, R.G.H. & Malloch, A.J.C. (2000) Modelling the potential effects of climate change on calcareous grasslands in Atlantic Europe. *Journal of Biogeography*, **27**, 347–358.
- Ellenberg, H. (1988) *Vegetation ecology of Central Europe*. Cambridge University Press, Cambridge.
- Forman, R.T.T. (1998) *Land mosaics: the ecology of landscapes and regions*. Cambridge University Press, Cambridge.
- 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.
- Frescino, T.S., Edwards, T.C. Jr & Moisen, G.G. (2001) Modeling spatially explicit forest structural attributes using Generalized Additive Models. *Journal of Vegetation Science*, **12**, 15–26.
- Gracia, C., Burriel, J.A., Ibàñez, J.J., Mata, T. & Vayreda, J. (2000a) *Inventari ecològic i forestal de Catalunya*. Regió Forestal V. CREAF, Bellaterra.
- Gracia, C., Burriel, J.A., Ibàñez, J.J., Mata, T. & Vayreda, J. (2000b) *Inventari ecològic i forestal de Catalunya*. Regió Forestal VI. CREAF, Bellaterra.
- Gracia, C., Burriel, J.A., Ibàñez, J.J., Mata, T. & Vayreda, J. (2000c) *Inventari ecològic i forestal de Catalunya*. Regió Forestal IV. CREAF, Bellaterra.
- Gracia, C., Burriel, J.A., Ibàñez, J.J., Mata, T. & Vayreda, J. (2001a) *Inventari ecològic i forestal de Catalunya*. Regió Forestal II. CREAF, Bellaterra.
- Gracia, C., Burriel, J.A., Ibàñez, J.J., Mata, T. & Vayreda, J. (2001b) *Inventari ecològic i forestal de Catalunya*. Regió Forestal III. CREAF, Bellaterra.

- Gracia, C., Ibáñez, J.J. & Rodà, F. (in press) Estat dels Boscos de Catalunya. 'Estat Del Medi Ambient a Catalunya', Generalitat de Catalunya (ed. by Departament de Medi Ambient de la Generalitat de Catalunya), DMA Generalitat de Catalunya, Barcelona.
- Guisan, A., Theurillat, J.-P. & Kienast, F. (1998) Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science*, **9**, 65–74.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.
- Hanley, J.A. & McNeil, B.J. (1983) A method of comparing the areas under receiver operating characteristic (ROC) curve. *Radiology*, **148**, 839–843.
- Harvey, L.E. (1996) *Third international conference/workshop on integrating GIS and environmental modelling, Santa Fe, NM Santa Barbara, CA*. National Center for Geographic Information and Analysis, Santa Barbara.
- Hastie, T.J. & Tibshirani, R. (1990) *Generalized additive models*. Chapman & Hall, London.
- Heikkinen, R.K. & Birks, H.J.B. (1996) Spatial and environmental components of variation in the distribution patterns of subarctic plant species at Kevo, N Finland — a case study at the meso-scale level. *Ecography*, **19**, 341–351.
- Huntley, B., Bartlein, P.J. & Prentice, I.C. (1989) Climatic control of the distribution and abundance of beech (*Fagus L.*) in Europe and North America. *Journal of Biogeography*, **16**, 551–560.
- Huntley, B., Berry, P.M., Cramer, W. & McDonald, A.P. (1995) Modelling present and potential future ranges of some European higher plants using climate response. *Journal of Biogeography*, **22**, 967–1001.
- Huston, M.A. (1994) *Biological diversity*. Cambridge University Press, Cambridge.
- Iverson, L.R. & Prasad, A. (1998) Predicting abundance for 80 tree species following climate change in the Eastern United States. *Ecological Monographs*, **68**, 465–485.
- Iverson, L.R., Prasad, A. & Schwartz, M.W. (1999) Modelling 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.
- Klaus, W. (1989) Mediterranean pines and their history. *Plant Systematics and Evolution*, **162**, 133–163.
- Korzukhin, M.D., TerMikaelian, M.T. & Wagner, R.G. (1996) Process versus empirical models: which approach for forest ecosystem management? *Canadian Journal of Forest Research*, **26**, 879–887.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Ninyerola, M., Pons, X. & Roure, J.M. (2000) A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *International Journal of Climatology*, **20**, 1823–1841.
- Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.
- Peix, J. (1999) *Programa de gestió del risc d'incendi forestal*. Generalitat de Catalunya, Barcelona.
- Peng, C. (2000) From static biogeographical model to dynamic global vegetation model: a global perspective on modelling vegetation dynamics. *Ecological Modelling*, **135**, 33–54.
- Pino, J., Vayreda, J. & Rodà, F. (2001) Distribución de especies arbóreas eurosiberianas en relación al tamaño del bosque. 'Montes Para la Sociedad Del Nuevo Lineo III Congreso Forestal Español', Vol. 3 (ed. by Consejería de Medio Ambiente de la Junta de Andalucía), pp. 202–207, Junta de Andalucía, Sevilla.
- Prentice, I.C., Bartlein, P.J. & Webb, T. (1991) Vegetation and climate change in eastern North America since the last glacial maximum. *Ecology*, **72**, 2038–2056.
- Prentice, I.C. & Helmisaari, H. (1991) Silvics of north European trees: compilation, comparisons and implications for forest succession modelling. *Forest Ecology and Management*, **42**, 79–93.
- Prentice, I.C. & Leemans, R. (1990) Pattern and process and the dynamics of forest structure: a simulation approach. *Journal of Ecology*, **78**, 340–355.
- Reed, R.A., Peet, R.K., Palmer, L.W. & White, P.S. (1993) Scale dependence on vegetation-environment correlations: a case study of a North Carolina piedmont woodland. *Journal of Vegetation Science*, **4**, 329–340.
- Rouget, M., Richardson, D.M., Lavorel, S., Vayreda, J., Gracia, C. & Milton, S.J. (2001) Determinants of distribution of six *Pinus* species in Catalonia, Spain. *Journal of Vegetation Science*, **12**, 491–502.
- Starfield, A.M. & Chapin, F.S. (1996) Model of transient changes in arctic and boreal vegetation in response to climate and land use change. *Ecological Applications*, **6**, 842–864.
- Swets, K.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.
- Sykes, M.T. (2001) Modelling the potential distribution and community dynamics of lodgepole pine (*Pinus contorta* Dougl. ex Loud.) in Scandinavia. *Forest Ecology and Management*, **141**, 69–84.
- Sykes, M.T. & Prentice, I.C. (1995) Boreal forest futures: modelling the controls on tree species range limits and transient responses to climate change. *Water, Air and Soil Pollution*, **82**, 415–428.
- Sykes, M.T. & Prentice, I.C. (1996) Climate change, tree species distributions and forest dynamics: a case study in the mixed conifer/northern hardwoods zone of northern Europe. *Climatic Change*, **34**, 161–177.
- Vayssières, M.P., Plant, R.E. & Allen-Diaz, B.H. (2000) Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, **11**, 679–694.
- Woodward, F.I. (1987) *Climate and plant distribution*. Cambridge University Press, Cambridge.

BIOSKETCHES

The authors of this paper participate in the European Union's ATEAM project.

Wilfried Thuiller and **Sandra Lavorel** are involved within ATEAM in the 'Biodiversity and nature conservation' sector. The main interest of the lead author is to study the determinants of species diversity at the biogeographical level in order to forecast the impacts of future global change on the distribution of diversity. He is currently working on a PhD examining patterns and causes of diversity distribution in Europe and South Africa.

Santi Sabate, Carles Gracia, Joan Pino and **Jordi Vayreda** contribute to the 'Forestry' sector of ATEAM.

ANNEXE 1.5

Do we need land-cover data to model species distributions in Europe?

Thuiller W., Araújo M.B., Lavorel S.

Journal of Biogeography (2004), sous presse



Do we need land cover data to model species distributions in Europe?

Wilfried Thuiller^{1,2*}, Miguel Araújo^{1,2} and Sandra Lavorel^{1,3}

¹Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de Recherche Scientifique, Montpellier cedex 5, France, ²Macroecology and Conservation Unit, University of Évora, Estrada dos Leões, Évora, Portugal and ³Laboratoire d'Ecologie Alpine, CNRS, Université J. Fournier, Grenoble Cedex 9, France.

ABSTRACT

Aim To assess the influence of land cover and climate on species distributions across Europe. To quantify the importance of land cover to describe and predict species distributions after using climate as the main driver.

Location The study area is Europe.

Methods (1) A multivariate analysis was applied to describe land cover distribution across Europe and assess if the land cover is determined by climate at large spatial scales. (2) To evaluate the importance of land cover to predict species distributions, we implemented a spatially explicit iterative procedure to predict species distributions of plants (2603 species), mammals (186 species), breeding birds (440 species), amphibian and reptiles (143 species). First, we ran bioclimatic models using stepwise generalized additive models using bioclimatic variables. Secondly, we carried out a regression of land cover (LC) variables against residuals from the bioclimatic models to select the most relevant LC variables. Finally, we produced mixed models including climatic variables and those LC variables selected as decreasing the residual of bioclimatic models. Then we compared the explanatory and predictive power of the pure bioclimatic against the mixed model.

Results (1) At the European coarse resolution, land cover is mainly driven by climate. Two bioclimatic axes representing a gradient of temperature and a gradient of precipitation explained most variation of land cover distribution. (2) The inclusion of land cover improved significantly the explanatory power of bioclimatic models and the most relevant variables across groups were those not explained or poorly explained by climate. However, the predictive power of bioclimatic model was not improved by the inclusion of LC variables in the iterative model selection process.

Main conclusion Climate is the major driver of both species and land cover distributions over Europe. Yet, LC variables that are not explained or weakly associated with climate (inland water, sea or arable land) are interesting to describe particular mammal, bird and tree distributions. However, the addition of LC variables to pure bioclimatic models does not improve their predictive accuracy.

Keywords

Bioclimatic models, climate gradients, land use effects, large-scale patterns, species distributions.

*Correspondence: Wilfried Thuiller, Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de Recherche Scientifique, 1919 route de Mende, 34293 Montpellier cedex 5, France. E-mail: wilfried.thuiller@cefe.cnrs-mop.fr

INTRODUCTION

It is widely accepted that distributions of plants and animals are broadly constrained by their physiological tolerances to climatic factors (Woodward, 1987, 1990). This generalization

is held to be true for species at a variety of spatial scales (Whittaker *et al.*, 2001), although there is a wide recognition that the importance of climate is best expressed at large spatial scales (Rahbek & Graves, 2001; Willis & Whittaker, 2002). Following this recognition, bioclimatic models are being used

to predict the distribution of plant and animal species at large spatial scales (for a review see Guisan & Zimmermann, 2000). However, bioclimatic models may produce inaccurate predictions when important local or regional factors are missing from input data in the models (Iverson *et al.*, 1999; Thuiller *et al.*, 2003b). As bioclimatic models assume species distributions to be at equilibrium with current environmental conditions, they are bound to over-predict distributions of species that are highly affected by human activities and that, consequently, may be missing from potentially suitable areas (Loehle & LeBlanc, 1996; Araújo, 2003). This problem should be particularly severe in regions where a positive relationship between species occurrence and human population density has been reported (e.g. Araújo, 2003).

For example, in Europe, land use changes, in particular agricultural intensification, have led to widespread changes in the distribution and abundance of many different taxa (Benton *et al.*, 2003). More specifically, numerous studies have revealed how important land use and other human activities affect mammal (Mladenoff *et al.*, 1999), bird (Chamberlain *et al.*, 1999), herptiles (Shenbrot *et al.*, 1991) and plant (Duckworth *et al.*, 2000) species distributions at regional scales.

As human management activities alter natural dynamics of species within ecosystems, bioclimatic models are expected to provide increasingly inefficient predictions of species distributions (e.g. Araújo *et al.*, 2001). A possibility to compensate for such human-induced factors is to include within-species distribution models additional variables expected to summarize important factors affecting local distributions of species. The inclusion of these variables should be expected to improve the accuracy of bioclimatic models (Loehle & LeBlanc, 1996; Midgley *et al.*, 2003). In this paper, we analyse the usefulness of land cover (LC) variables to predict plant, bird, mammal and herptile species distributions over Europe. We also distinguish between the explanatory and predictive power of land cover. Variables with high explanatory value help to understand the determinants of species distribution, whereas variables with predictive value play an important role for model development without necessarily being ecologically relevant. Some variables can play both roles, whereas others are more orientated toward one role. The distinction between both components has thus serious implication in practice.

We address the following questions:

1. How much of land cover distribution is explained by climate?
2. How does land cover increase the explanatory power of bioclimatic models at large spatial scales?
3. How does land cover increase the predictive power of bioclimatic models at large spatial scales?

METHODS

Species data

Original species data included 2294 plant (Jalas & Suominen, 1972–1996), 186 mammal (Mitchell-Jones *et al.*, 1999), 143 amphibian and reptile (Gasc *et al.*, 1997) and 440 breeding

bird species (Hagemeyer & Blair, 1997). Terrestrial vertebrates include all known species (Araújo *et al.*, 2001) whereas plants comprise only ~ 20% of the European flora (Humphries *et al.*, 1999). The grid used is based on the *Atlas Florae Europaeae* (AFE; Lahti & Lampinen, 1999), with cells boundaries typically following the 50 km lines of the Universal Transvers Mercator (UTM) grid. The remaining atlases use slightly different grid-systems, including different rules to represent data on islands and coasts. Hence, vertebrate data were converted to the AFE grid system by identifying unique correspondence between cells in these grids (Williams *et al.*, 2000). The mapped area (2434 grid cells) includes western, northern and southern Europe, but exclude most of the eastern European countries where recording effort was both less uniform and less intensive (Williams *et al.*, 2000).

As plants are by far the largest including a great variety of responses, we divided them into three life-form-based groups according to Tutin *et al.* (1964–1993): herbs, shrubs and trees.

Bioclimatic data

We used a comprehensive set of bioclimatic variables for analyses (T.D. Mitchell, 2002, unpublished data): mean annual, winter and summer precipitation, mean annual temperature and minimum temperature of the coldest month, growing degree days (> 5°) and an index of humidity (mean ratio of annual actual evapotranspiration over annual potential evapotranspiration). Mean values are averaged from 1960 to 1990. These data were supplied on a 10' grid, covering Europe and then aggregated by averaging to 50 × 50 km UTM in order to match with the resolution of species data.

Land cover data

The land cover data were developed at a spatial resolution of 10' for grid cells based on the ATEAM project geographical window (<http://www.pik-potsdam.de/ateam/>). The 10' grid cells were derived by aggregation from the PELCOM land cover database. PELCOM is a 1-km pan European land cover database developed mainly from remotely sensed data. The classification methodology is based on a regional and integrated approach of the NOAA-AVHRR satellite data and ancillary information such as topographic features (Mücher, 2000). Although finer spatial resolution databases exist such as CORINE (CEC, 1993), PELCOM was selected due to its complete spatial coverage of the European window, and because of the homogeneity of the methodology used for the land cover classification. PELCOM is also the most up-to-date of pan-European land cover databases.

The percentages of each land cover class were calculated for the individual 10' grid cells from the 1 km PELCOM cells. The data comprised four classes of land cover: forest, agriculture, urban (perurban) and others. Forests were further subdivided into three classes: percentage of deciduous, coniferous and mixed forest (respectively named perdec, percon and permixf). Agriculture was also divided into three

classes: the percentage of arable lands, grasslands and permanent crops (respectively named perara, pergrass and prepermc). The 'other' land cover class comprised: percentage of semi-natural areas (divided into areas of shrubland, and barrenland, respectively named pershrub and perbarrenl), inland waters (perinwater), wetlands (perwetland), permanent ice and snow (perpice) and sea (persea). Similarly to climate, we aggregated 10' grid data in 50 × 50-km grid to match with species data by calculating the percentage of each class in the new grid system.

Model

Bioclimatic distribution of land cover in Europe

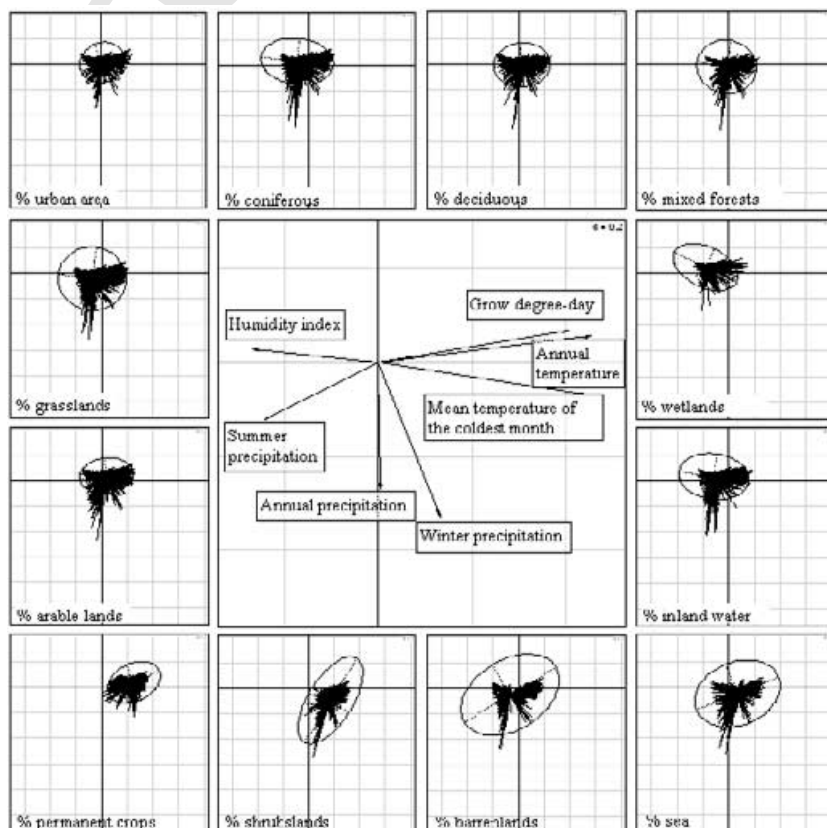
Given that climate governs global patterns of land cover (Dale, 1997), land cover and climate are not fully independent. Hence, before implementing predictive models of species distributions we investigated patterns of co-variation between land cover and climate variables. This was performed with the Outlying Mean Index (OMI), a co-inertia analysis recently developed by Doledec *et al.* (2000). This analysis is similar to the well-know canonical correspondence analysis (CCA: ter Braak, 1986), but allows to separate land cover classes and to measure the distance between the mean bioclimatic conditions used by each land cover class and the mean bioclimatic conditions across Europe. Moreover, unlike CCA, this method has the advantage of making no assump-

tions about the shape of land cover response curves to the bioclimate. We applied OMI to separate land cover classes in Europe by performing a normalized PCA on the bioclimatic table and linking the land cover table to the PCA output table using the adequate diagonal and metric matrices (Thuiller *et al.*, 2003c)

The first two selected axes explained 96% of the total inertia, or variation in the data. The first axis was mainly related to temperature and the second to precipitation (Fig. 1). Projecting land cover (LC) variables on the two-dimensional bioclimatic space allows describing their distributions (Fig. 1). For instance, permanent crops occurred mainly at cooler temperatures, whereas grasslands were mainly located in mid-altitude mountains less intensively disturbed by human management. However, some LC variables were found to be widespread within climatic space. This was the case, for example, of percentage of sea, inland water, arable lands, urban area or coniferous forests (Fig. 1).

Given these patterns of co-variation between land cover and bioclimatic variables, if we modelled species distribution with both kinds of variables and with a stepwise variable-selection procedure (Chambers & Hastie, 1997) we would rarely select LC variables as bioclimatic variables explains the most important components of their distributions. However, LC variables may still contribute significantly to model species distribution in Europe when more regional patterns of LC, corresponding to the residual part of the OMI, may influence distributions of particular species.

LOW RESOLUTION FIG



8 Figure 1 Centre: Canonical weight of bioclimatic variables and relationships with the first two axes of the OMI. Sides: Distribution of land cover variables on the first two axes of the OMI analysis. Crosses identify the position of the land cover variables in the bioclimatic space using the canonical weights of bioclimatic variables. Ellipses represent the land cover inertia representing at least 90% of points for each variable. The dashed lines emanating from the centroids of each ellipse represent the major and minor axes of the ellipse.

Rule of procedure	
Step 1	Stepwise GAM using only bioclimatic variables (GAM-clim)
Step 2	Stepwise GAM on residuals from step 1 using only land cover variables
Step 3	GAM using bioclimate variables retained in step 1 and LC variables retained in step 2 (GAM-clim-LC)
Step 4	Evaluation of GAM-clim accuracy on evaluation data using ROC curve
Step 5	Comparative ANOVA between GAM-clim and GAM-clim-lu (test)
Step 6	If test is significant: evaluation of GAM-clim-lu on evaluation data using ROC curve Else, evaluation of GAM-clim on evaluation data (same as step 3)
Step 7	Averaging of ROC curve from GAM-clim on all species
Step 8	Averaging of ROC curve from GAM-clim-lu (if test is significant) and GAM-clim (if test is not significant)

Table 1 Descriptive steps of the iterative procedure.

Including LC variables into bioclimatic models.

In order to explore for the potential residual contribution of land cover explaining current regional distributions of species, we implemented an iterative framework (Table 1) based on a particular implementation of Generalized Additive Models (GAM). Here we first run stepwise GAM with bioclimatic variables and then assessed if the inclusion of LC variables improved the explanatory and predictive performance of bioclimatic models.

Original data were randomly divided into a calibration (70%) and an evaluation (30%) data set. Species with more than 10 occurrences in the calibration data were selected for model calculations (165 mammals, 117 herptiles, 387 breeding birds and 1527 plants remaining). Stepwise GAMs were then run using selected bioclimatic variables. Then we ran stepwise GAM of residuals against LC variables. Selected LC variables for each species were then added to the set of bioclimatic variables to produce mixed bioclimatic models including land cover. To assess the explanatory power of land cover, comparative ANOVAs were performed between the pure bioclimatic and the mixed models. To assess the importance of land cover as predictive variables a test of accuracy was performed on the evaluation data using the relative operating characteristic (ROC) curve (Pearce & Ferrier, 2000; Thuiller *et al.*, 2003a). The ROC curve was applied to pure bioclimatic models and to the mixed models. Subjective guidelines (Swets, 1988) suggest very good agreement for AUC above 0.9 (Table 1). All analyses were performed with Splus (Anon., 1999).

RESULTS

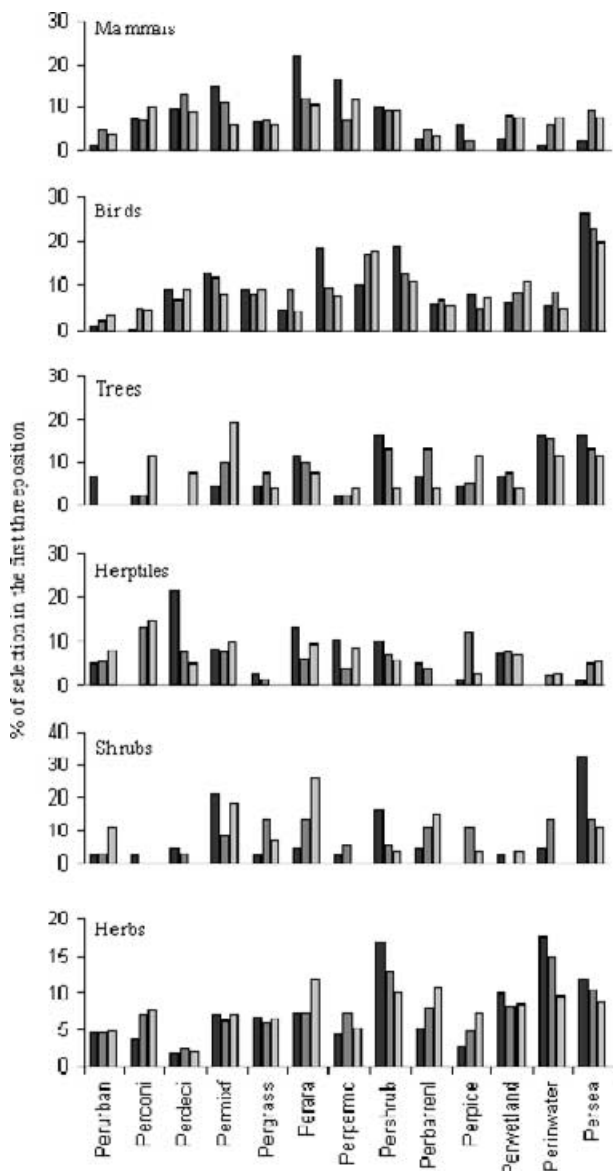
How does land cover improve the explanatory power of bioclimatic models at a large spatial scale?

The inclusion of land cover increased the explanatory power of bioclimatic models for species of all groups (Table 2). This was particularly true for mammals, birds and trees, where the inclusion of land cover increased the explained deviance of bioclimatic models by more than 60%.

Table 2 Explanatory power of land cover variables. Number of species corresponds to the number of species studied in this analysis. Step 2 corresponds to the number of species where land cover variables were retained as decreasing residuals of bioclimatic models (step 2, Table 1). Step 5 corresponds to the number of species for which mixed models had a higher explained deviance than pure bioclimatic models (step 5, Table 1). Percentage corresponds to the percentage of cases where the inclusion of land cover increased the explanatory power of bioclimatic models.

	No. of species	Step 2	Step 5	Percentage
Mammals	165	162	107	65
Birds	387	377	241	62
Trees	71	71	43	61
Shurbs	78	78	43	55
Herptiles	117	110	62	53
Herbs	1378	1265	624	45

In order to explore the additional effects of land cover after bioclimatic models have been adjusted, we selected a few species with established relationships with land cover at lower spatial scales. This was the case for wolf *Canis lupus*, whose populations are known to be inversely correlated with human population and activities (Breitenmoser, 1998; Massolo & Meriggi, 1998). In our analysis, the LC variables retained for this species were percentage of urban areas and percentage of grasslands. The response curves to these two variables were exponential negative and linearly negative, respectively, showing that even at large spatial scales important LC variables can be detected by models. Overall, the three LC variables most often retained by models for mammals were the percentages of arable land, permanent crops and mixed forests (Fig. 2, mammals). For birds, we analysed models for the grey partridge *Perdix perdix*. Robinson *et al.* (2001) showed that the grey partridge has declined in Britain as a consequence of the intensification of agricultural practices and the decrease of available arable land. In support of this study, our analysis selected arable land as the most important LC variable for the species (Table 1). More generally, the inclusion of the percentage of arable land into models of distributions for



8 Figure 2 Percentage of models using each land cover variable as one of the first three land cover variables selected in step 2 for decreasing the residuals of pure bioclimatic models. Dark grey, grey and clear greys bars represent, respectively, the most, the second most and the third most explanatory selected variables during the process for each group of taxa.

farmland birds increased significantly the explanatory power of models. As a whole, agriculture (percentage of arable land and permanent crops) and presence of sea were the most important LC variables accounting for residuals in bird distributions (Fig. 2, birds). For trees, we focused on *Quercus petraea*, a Euro-Siberian species with wide distribution (Tutin *et al.*, 1964–1993). The only selected LC variable was the percentage of deciduous forest, with a strong positive relationship. This is, indeed, a deciduous tree dominant in central European deciduous forests. Another example is *Q. pedunculiflora*, which was negatively correlated to permanent crops and positively related to the percentage of shrublands. For trees, in general,

there were no unequivocally dominant LC variables explaining residuals, although presence of sea, inland water and shrubs appeared often as useful to explain distributions (Fig. 2, plants–trees).

For herptiles and shrubs, the inclusion of land cover increased the explanatory power of bioclimatic models for more than half of the species (Table 2). Herptiles responded mainly to percentage of deciduous, coniferous forests and arable land (Fig. 2, herptiles), while shrubs were more related to percentage of sea, arable land and mixed forests (Fig. 2, plants–shrubs). These observations are consistent with other published records. For instance, Martín & Lopez (2002) showed that the lizard *Lacerta lepida* occurred mainly under *Quercus* forest and understorey low bushes and avoided croplands. Here, the two retained variables were percentage of mixed forest (positive relation) and permanent crops (negative relation). For shrubs, we chose to investigate the response of the pioneer white willow *Salix alba* and found that the percentage of arable land (positive relationship) was the only additional variable selected for the species.

Finally, herbs were the group for which the inclusion of land cover produced the least improvement over pure bioclimatic models (Table 2). The most relevant variables selected were the percentage of inland water, sea and shrubs (Fig. 2, plants–herbs). We focused on the woodland herbs group, which are mainly correlated to percentage of deciduous and mixed forest. For instance, the yellow anemone *Anemone ranunculoides* that typically occurs under beech forests (Tutin *et al.*, 1964–1993) has a positive unimodal relationship with percentage of deciduous and a negative relationship with mixed forests and permanent crops. The superb pink *Dianthus superbus*, which occurs principally under open forests (Tutin *et al.*, 1964–1993), was as expected positively related to mixed forests and negatively to arable land.

How does land cover improve the predictive power of bioclimatic models at a large spatial scale?

The predictive power of models was assessed using AUC on evaluation data. We concluded that models, on average, provided very good results for all groups (Table 3).

Shrubs and herptiles were the species best predicted by models with an average AUC of 0.95 for evaluation. Conversely, mammals had the lowest (although still rather high) mean AUC values with 0.914. The lowest values of AUC were generally comprised 0.66 and 0.80 while maximum values of AUC were equal or very close to 1 (Table 3).

Predictions using bioclimatic and LC data showed very similar results to predictions using bioclimatic data alone. Although there were no significant differences between AUC for both analyses, some general trends can be outlined. First, the overall minimum AUC was generally higher (but not for trees) for models using LC data than for models that did not use it (Table 3). Secondly, standard deviations of AUC for models using bioclimatic and LC data were lower than for models using only bioclimatic variables (Table 3). The last two

Table 3 Statistical summary of AUC from the ROC curve procedure displaying the standard deviation (SD), the minimum (min), the average (mean) and maximum (max) for each group according to the models (bioclimatic model against mixed including land cover variables). Difference correspond to the net difference between mean AUC from bioclimatic and mixed models.

	Bioclimatic model				Mixed model				Difference
	SD	Min	Mean	Max	SD	Min	Mean	Max	
Herptiles									
Cali	0.0467	0.8017	0.9573	0.9999	0.0413	0.8099	0.9627	0.9999	-0.0054
Eval	0.0505	0.7791	0.9497	0.9988	0.0487	0.7933	0.9504	0.9988	-0.0007
Mammals									
Cali	0.0582	0.7457	0.9277	0.9992	0.0522	0.7812	0.9358	0.9992	-0.0081
Eval	0.0632	0.7267	0.9140	0.9987	0.0603	0.7552	0.9133	0.9975	0.0007
Birds									
Cali	0.0459	0.7853	0.9388	0.9990	0.0421	0.8227	0.9447	0.9990	-0.0059
Eval	0.0525	0.7592	0.9254	0.9980	0.0512	0.7592	0.9260	0.9980	-0.0006
Herbs									
Cali	0.0419	0.7341	0.9568	1.0000	0.0419	0.7341	0.9567	1.0000	0.0001
Eval	0.0503	0.6628	0.9313	1.0000	0.0503	0.6628	0.9313	1.0000	0.0000
Shrubs									
Cali	0.0314	0.8692	0.9708	1.0000	0.0256	0.8950	0.9743	1.0000	-0.0035
Eval	0.0421	0.8041	0.9530	1.0000	0.0406	0.8046	0.9558	1.0000	-0.0028
Trees									
Cali	0.0277	0.8950	0.9642	0.9993	0.0327	0.8482	0.9615	0.9993	0.0027
Eval	0.0347	0.8188	0.9436	1.0000	0.0384	0.8136	0.9414	1.0000	0.0032

points support the idea that the inclusion of LC variables may help improve predictions for species where bioclimatic models yielded poorer predictions.

Amongst the examples provided, there is no species for which the inclusion of land cover increased significantly the predictive accuracy of bioclimatic models. However, land cover improved bioclimatic models (on evaluation data) only for a limited number of species: 34% of mammals, 40% of herptiles, 34% of birds and 21% of plants. Animals appeared to be more related to land cover, but differences between AUC on evaluation data are too narrow to draw clear conclusions on the benefit from including land cover to predict animal species distributions at large scales.

DISCUSSION

Distribution of land cover in Europe

The OMI multivariate analysis of environmental data revealed that, over Europe, the spatial variation of land cover is highly correlated to spatial variation among bioclimatic variables. This is an important result given that it explains why including both types of information into species distribution models may bring redundant information. The degree to which this is the consequence of the large spatial scale and coarse resolution of data remains unknown, but it is possible that models of species distributions using a smaller spatial extent and using finer resolution data might produce a different pattern (Franklin,

1995; Collingham *et al.*, 2000). Indeed, a hierarchical scheme of environmental controls on species distributions has been suggested (but not demonstrated), in which climatic variables are large-scale determinants, followed by geology, land cover and topography, which moderate many of the effects of macroclimatic variables (Franklin, 1995). In our analysis, although a large proportion of the variability of land cover distribution was described by climate, there were residuals related to regional land cover patterns that could not be explained by climate. There were indeed some LC variables either weakly or not related to climatic factors, such as inland water, sea, arable land and coniferous forests. Sea is present around a large proportion of Europe and is obviously not dependent on climate. Similarly to sea, in Europe, inland water is not restricted by climate and is present across all regions. More surprisingly the distribution of coniferous forests did not appear to be influenced by climate. The reason for this is that this class brings together a variety of species with a large range of climatic and ecological requirements, which makes it difficult to find a general pattern (Richardson, 1998). For instance, the *Pinus* genus is present in the Siberian plains (*Pinus sibirica*), in the Alps mountains (*P. cembra*), in restricted areas of the Mediterranean region (*P. halepensis*) and in the Euro-Siberian region (*P. sylvestris*). Such a variety of chorotypes explains why the conifer distributions do not appear to be controlled by climate in our analysis, although the species making-up this class are indeed related to climate at European scale (Richardson, 1998). Arable lands occur across

the whole of Europe mainly because this class embeds adequate lands suitable for agriculture independently of the type of cultures.

These widespread LC variables that are little related, or even unrelated, to climate at the European scale are, however, influenced by multiple regional or local factors such as soil type, slope, aspect or groundwater distribution and availability, that were not taken into account in our analysis.

Explanation of current species distributions

As widespread LC variables are not restricted by climate, it was not surprising to find that they were often retained by GAM models (Table 1, step 2), and decreased significantly the residual from pure bioclimatic models for many species. Arable lands and permanent crops improved significantly the explanatory power of bioclimatic models for several breeding birds and mammals. Our results corroborate previous studies showing the importance, positive or negative, of arable land for birds, mammals and plants from regional to local scales (e.g. Robinson *et al.*, 2001; Robinson & Sutherland, 2002; Benton *et al.*, 2003). These authors showed that the intensification of agriculture and the contraction of arable cultivations have produced important local extinctions of birds and mammals (Robinson *et al.*, 2001). Our study emphasized that even at large spatial scales and coarse resolutions, human-related variables could be useful to describe and explain species distributions (e.g. *Perdix perdix*).

However, our approach is only correlational and the ecological meaning of the observed relationships between land cover and species distributions is not always obvious. Although bioclimatic models identify correlational relationships, there have been a number of studies investigating causal relationships between species and bioclimatic variables (at least for plants) (Woodward, 1987, 1992). This kind of analysis has not been developed, as far as we are aware, for species and land cover at large spatial scales. The causal relationships between variables such as land cover and species distributions are not always easy to uncover since they are often indirect. This is the case for plants as land cover can be seen more as a limiting factor (at least for human-oriented variables) than a factor having direct physiological impact. However, for vertebrates the relationship can be more direct. For instance, high quality habitats for winter farmland birds (stubble, game cover and game feeders) occur principally in arable lands (Benton *et al.*, 2003). Granivorous passerines and small rodents are obviously strongly related to permanent crops and arable land where they find refuges and food.

Prediction of species distribution ranges

From a predictive standpoint, LC variables did not improve significantly the predictive performance of our models built on evaluation data. Several non-exclusive hypotheses can be proposed to explain this pattern:

(1) Goodness-of-fit does not make an assessment of predictive performance (although their failure underscores aspects of the relationship between the modelled species distributions and the explanatory variables used in the model) (Hosmer & Lemeshow, 1989). Even if some LC variables improve the explanatory power of bioclimatic models, this does not imply that they would also improve the predictive performance.

(2) Climate explains almost all LC variation and the residuals are not large enough to improve the predictive performance of bioclimatic models. The residuals of the climate–LC relationships may have a strongly regional distribution and therefore do not provide a strong underlying gradient affecting species distributions. Hence they may not be relevant for predictive purposes.

(3) In modelling species against bioclimatic variables it is often assumed that variables can be grouped into three important groups: direct, resources and indirect variables (Austin & Smith, 1989). Direct or resource gradients provide mechanistic and physiological explanations for distributions of species. Models that use such variables are supposed to have greater predictive performance than models using indirect variables, hence they should be more accurate when predicting distributions over large areas or at other times in the future (Guisan & Zimmermann, 2000). Indirect variables, such as land cover, are supposed to have little direct physiological relevance for species. Hence they should only be applied within a limited geographical extent without significant errors, because in a different region, or time, the same LC variables can correspond to a different combination of direct and resource gradients (Guisan & Zimmermann, 2000).

(4) Using the 50 × 50-km grid cell resolution across Europe, we demonstrated that LC distribution was driven mainly by climatic gradients. However, inside each grid cell there is an underlying high heterogeneity of LC that is lost (see the PELCOM database), while this heterogeneity is a crucial factor for predicting current species occurrences (Cowling & Lombard, 2002; Benton *et al.*, 2003). This constrain by the data combines two potential problems. First, the problem of using data with insufficient resolution to explain observed patterns (Robinson *et al.*, 1992; Pearson, 1993). Secondly, the possibility that LC variables used may not include all relevant factors. For example, variables representing the indices of agricultural intensification (Mader, 1984), habitat heterogeneity (Bascompte & Solé, 1996; Allen & O'Connor, 2000; Gaston *et al.*, 2002), frequency and intensity of perturbation (Croonquist & Brooks, 1993), density of roads (Mader, 1984) could be more relevant to improve the predictive ability of bioclimatic models for some species, even at a 50 × 50-km resolution.

ACKNOWLEDGMENTS

We thank the national atlas representatives and the many volunteer fieldworkers who contributed to the atlas records, species distribution data was kindly supplied by J.P. Gasc (Societas Europaea Herpetologica), W.J.M. Hagemerijer

(European Bird Census Council), Raino Lampinen (Atlas Flora Europaeae), and A.J. Mitchell-Jones (Societas Europaeae Mamalogica). Tim Mitchell (Tyndall Centre for Climate Research) generated climate data, Markus Erhard and Sönke Zaehle ('Potsdam Institut für Klimafolgenforschung') aggregated climate surfaces to the AFE grid. Marc Rounsevell and Isabelle Reginster (University of Louvain-La-Neuve) for aggregated PELCOM land cover data to the AFE grid. Current research is funded by the European Commission ATEAM (EVK2-CT-2000-00075) project, MBA also thanks FCT (SPRH/BPD/5547/2001) for funding.

REFERENCES

- Allen, A.P. & O'Connor, R.J. (2000) Interactive effects of land cover and other factors on regional bird distributions. *Journal of Biogeography*, **27**, 889–900.
- Anon. (1999) *S-PLUS 2000 user's guide*. MathSoft, Seattle, WA.
- Araújo, M.B. (2003) The coincidence of people and biodiversity in Europe. *Global Ecology and Biogeography*, **12**, 5–12.
- Araújo, M.B., Humphries, C.J., Densham, P.J., Lampinen, R., Hagemeyer, W.J.M., Mitchell-Jones, A.J. & Gasc, J.P. (2001) Would environmental diversity be a good surrogate for species diversity? *Ecography*, **24**, 103–110.
- Austin, M.P. & Smith, T.M. (1989) A new model for the continuum concept. *Vegetatio*, **83**, 35–47.
- Bascompte, J. & Solé, R.V. (1996) Habitat fragmentation and extinction thresholds in spatially explicit models. *Journal of Animal Ecology*, **65**, 465–473.
- Benton, T.G., Vickery, J.A. & Wilson, J.D. (2003) Farmland biodiversity: is habitat heterogeneity the key? *TREE*, **18**, 182–188.
- ter Braak, C.J.F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, 1167–1179.
- Breitenmoser, U. (1998) Large predators in the Alps: the fall and rise of man's competitors. *Biological Conservation*, **83**, 279–289.
- CEC (1993) *CORINE land cover technical guide*, European Union. Directorate-General Environment, Nuclear Safety and Civil Protection. Office for Official Publications of the European Communities, Luxembourg.
- Chamberlain, D.E., Wilson, A.M., Browne, S.J. & Vickery, J.A. (1999) Effects of habitat type and management on the abundance of skylarks in the breeding season. *Journal of Applied Ecology*, **36**, 856–870.
- Chambers, J.M. & Hastie, T.J. (1997) *Statistical models in* **2** S. Chapman & Hall, London.
- Collingham, Y.C., Wadsworth, R.A., Huntley, B. & Hulme, P.E. (2000) Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology*, **37**, 13–27.
- Cowling, R.M. & Lombard, A.T. (2002) Heterogeneity, speciation/extinction history and climate: explaining regional plant diversity patterns in the Cape Floristic Region. *Diversity and Distributions*, **8**, 163–179.
- Croonquist, M.J. & Brooks, R.P. (1993) Effects of habitat disturbance on bird communities in riparian corridors. *Journal of Soil and Water Conservation*, **48**, 65–70.
- Dale, V.H. (1997) The relationship between land-use change and climate change. *Ecological Applications*, **7**, 753–769.
- Doledec, S., Chessel, D. & Gimaret-Carpentier, C. (2000) Niche separation in community analysis: a new method. *Ecology*, **81**, 2914–2927.
- Duckworth, J.C., Bunce, R.G.H. & Malloch, A.J.C. (2000) Vegetation-environment relationships in Atlantic European calcareous grasslands. *Journal of Vegetation Science*, **11**, 15–22.
- Franklin, J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, 474–499.
- Gasc, J.-P., Cabela, A., Crnobrnja-Isailovic, J., Dolmen, D., Grossenbacher, K., Haffner, P., Lescure, J., Martens, H., Martinez Rica, J.P., Maurin, H., Oliveira, M.E., Sofianidou, T.S., Veith, M. & Zuiderwijk, A. (1997) *Atlas of amphibians and reptiles in Europe*. Societas Europaea Herpetologica & Museum National d'Histoire Naturelle, Paris.
- Gaston, K.J., Pressey, R.L. & Margules, C.R. (2002) Persistence and vulnerability: retaining biodiversity in the landscape and in protected areas. *Journal of Biosciences*, **27**, 361–384.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Hagemeyer, W.J.M. & Blair, M.J. (1997) *The EBCC atlas of European breeding birds, their distribution and abundance*. Poyser, London.
- Hosmer, D.W. & Lemeshow, S. (1989) *Applied logistic regression*. Wiley, New York.
- Humphries, C., Araújo, M.B., Williams, P., Lampinen, R., Lahti, T. & Uotila, P. (1999) Plant diversity in Europe: Atlas Florae Europaeae and WORLDMAP. *Acta Botanica Fennica*, **162**, 11–21.
- Iverson, L.R., Prasad, A. & Schwartz, M.W. (1999) Modelling potential future individual tree-species distributions in the Eastern United States under climate change scenario: a case study with *Pinus virginiana*. *Ecological Modelling*, **115**, 77–93.
- Jalas, J. & Suominen, J. (1972–1996) *Atlas Florae Europaeae*. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.
- Lahti, T. & Lampinen, R. (1999) From dot maps to bitmaps – Atlas Florae Europaeae goes digital. *Acta Botanica Fennica*, **162**, 5–9.
- Loehle, C. & LeBlanc, D. (1996) Model-based assessments of climate change effects on forests: a critical review. *Ecological Modelling*, **90**, 1–31.
- Mader, H.J. (1984) Animal habitat isolation by roads and agricultural fields. *Biological Conservation*, **29**, 81–96.
- Martin, J. & Lopez, P. (2002) The effect of Mediterranean dehesa management on lizard distribution and conservation. **3** *Biological Conservation*, **108**, 231–239.
- Massolo, A. & Meriggi, A. (1998) Factors affecting habitat occupancy by wolves in northern Apennines (northern Italy): a model of habitat suitability. *Ecography*, **21**, 97–107.

- Midgley, G.F., Hannah, L., Millar, D., Thuiller, W. & Booth, A. (2003) Developing regional and species-level assessments of climate change impacts on biodiversity: a preliminary study in the Cape Floristic Region. *Biological Conservation*, **112**, 87–97.
- Mitchell-Jones, A.J., Amori, G., Bogdanowicz, W., Krystufek, B., Reijnders, P.J.H., Spitzenberger, F., Stubbe, M., Thissen, J.M.B., Vohralik, V. & Zima, J. (1999) *Atlas of European mammals*. Academic Press, London.
- Mladenoff, D.J., Sickley, T.A. & Wydeven, A. (1999) Predicting gray wolf landscape recolonization: logistic regression models vs. new field data. *Ecological Applications*, **9**, 37–44.
- Mücher, C.A. (ed.) (2000) *PELCOM project 'development of a consistent methodology to derive land cover information on a European scale from remote sensing for environmental modelling'*. Final report, February 2000, Part 2, EU DG XII-D.
- Pearce, J. & Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127–147.
- Pearson, S.M. (1993) The spatial extent and relative influence of landscape-level factors on wintering bird populations. *Landscape Ecology*, **8**, 3–18.
- Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4534–4539.
- Richardson, D.M. (1998) *Ecology and biogeography of Pinus*. Cambridge University Press, Cambridge.
- Robinson, R.A. & Sutherland, W.J. (2002) Post-war changes in arable farming and biodiversity in Great Britain. *Journal of Applied Ecology*, **39**, 157–176.
- Robinson, G.R., Holt, R.D., Gaines, M.S., Hamburg, S.P., Johnson, M.L., Fitch, H.S. & Martinko, E.A. (1992) Diverse and contrasting effects of habitat fragmentation. *Science*, **257**, 524–526.
- Robinson, R.A., Wilson, J.D. & Crick, H.Q.P. (2001) The importance of arable habitat for farmland birds in grassland landscapes. *Journal of Applied Ecology*, **38**, 1059–1069.
- Shenbrot, I.G., Rogovin, K.A. & Surov, A.V. (1991) Comparative analysis of spatial organization of desert lizard communities in Middle Asia and Mexico. *Oikos*, **61**, 157–168.
- Swets, K.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.
- Thuiller, W., Araújo, M.B. & Lavorel, S. (2003a) Generalized models versus classification tree analysis: a comparative study for predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, **14**, 669–680.
- Thuiller, W., Vaydera, J., Pino, J., Sabaté, S., Lavorel, S. & Gracia, C. (2003b) Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology and Biogeography*, **12**, 313–325.
- Thuiller, W., Lavorel, S., Midgley, G.F., Lavergne, S. & Rebelo, A.G. (2003c) Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* species in the Cape Floristic Region. *Ecology* (in press).
- Tutin, T.G., Burges, N.A., Chater, A.O., Edmondson, J.R., Heywood, V.H., Moore, D.M., Valentine, D.H., Walters, S.M. & Webb, D.A. (1964–1993) *Flora Europaea*. Cambridge University Press, Cambridge.
- Whittaker, R.J., Willis, K.J. & Field, R. (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, **28**, 453–470.
- Williams, P., Humphries, C., Araújo, M.B., Lampinen, R., Hagemeyer, W., Gasc, J.-P. & Mitchell-Jones, T. (2000) Endemism and important areas for representing European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belgian Journal of Entomology*, **2**, 21–46.
- Willis, K.J. & Whittaker, R.J. (2002) Species diversity-scale matters. *Science*, **295**, 1245–1248.
- Woodward, F.I. (1987) *Climate and plant distribution*. Cambridge University Press, Cambridge.
- Woodward, F.I. (1990) The impact of low temperatures in controlling the geographical distribution of plants. *Philosophical Transactions of the Royal Society of London B*, **326**, 593–595.
- Woodward, F.I. (1992) Predicting plant responses to global environmental change. *New Phytologist*, **122**, 239–251.

BIOSKETCHES

- 7 The authors of this paper participated in the European Union's ATEAM project. **Wilfried Thuiller**, **Miguel Araújo** and **Sandra Lavorel** are involved within ATEAM in the 'Biodiversity and nature conservation' sector. The main interest of the lead author is to study the determinants of species diversity at the biogeographical level in order to forecast the impacts of global change future on distributions of diversity. He is currently working on a PhD examining patterns and causes of diversity distribution in Europe and South Africa.

ANNEXE 1.6

Downscaling European species atlas distributions to a finer resolution: implications for conservation planning

Araújo M.B., Thuiller W., Williams P.H., Reginster I.

Global Ecology & Biogeography (2004) sous presse

Downscaling European species atlas distributions to a finer resolution: implications for conservation planning

Miguel B. Araújo^{1,2}, Wilfried Thuiller^{1,2}, Paul H. Williams³, Isabelle Reginster⁴

¹Macroecology and Conservation Unit, University of Évora, Portugal; ²Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, Montpellier, France; ³Biogeography and Conservation Laboratory, The Natural History Museum, London, UK; ⁴Département de Géographie, Université Catholique de Louvain, Louvain-la-Neuve, Belgique

Running title: downscaling species atlas distributions

Word count: 4147, including references

Address for correspondence: Email: mba@uevora.pt

ABSTRACT

Aim: One of the limitations to using species atlas distributions in conservation planning is their coarse resolution, relative to the needs of local planners. We outline a simple approach downscale original coarse-scale species atlas distributions to a finer resolution. If such procedure yielded accurate downscaled predictions, then it could be an aid to using available distribution-atlas data in real-world local decisions.

Location: Europe

Methods: An iterative approach based on generalised additive modelling is outlined. Results are assessed in relation to models' ability to reproduce 1) spatial patterns of coincidence in species richness scores among different groups; and 2) spatial patterns of coincidence in richness, rarity and complementarity hotspots.

Results: Although hotspots were not fully coincident between non-modelled and modelled data and between modelled non-downscaled and modelled downscaled data, there was evidence that downscaled distributions were able to recover original patterns of coincidence between richness, at least for terrestrial vertebrate groups. Downscaled distributions were also able to recover morphological and climatic gradients originally blurred by coarse-resolution data.

Main Conclusions: Despite uncertainties, downscaling procedures might help identify reserves that are more meaningfully related to local patterns of environmental variation.

KEYWORDS: *distribution atlas; downscaling; generalised additive models; species modelling; reserve selection.*

INTRODUCTION

Distribution atlases provide a convenient representation of species distributions in near-equal-area grid cells on maps. Such representations have been widely used to study species distribution patterns of abundance, rarity, richness, turnover and assemblage composition at varying spatial scales (e.g. Araújo 2003; Koleff & Gaston 2002; Turner et al. 1988; van Rensburg et al. 2002; Williams et al. 1999). Atlas data have also been used to investigate and test the performance of different conservation-planning methodologies, such as reserve selection (Araújo et al. 2001; Araújo & Williams 2000; Araújo et al. 2002a,b; Rebelo & Siegfried 1992). However, there is a concern that the scale of most atlases is too coarse for real-world conservation planning applications (Hopkinson et al. 2000; Stoms 1992; Stoms 1994; Rouget 2003). In response to this problem there have been attempts to downscale species distributions to finer resolutions by combining remotely sensed data and expert opinion to assign species to habitats or land-cover classes thought to be suitable for species (e.g. Boitani et al. 1999;

Wisiz, in preparation). One of the problems with this approach is that there may be insufficient expert knowledge of species-habitat relationships for many species and areas (e.g. Freitag et al. 1998). In such cases the usefulness of expert-driven approaches may be limited to few well-known, often charismatic, species.

An alternative is to use empirical modelling techniques that explore the correlation between a response (e.g. species' occurrence records) and sets of predictor (e.g. climate and land cover) variables to extrapolate distributions into another spatial resolution (e.g. Barbosa et al. 2003). Here we provide an example of such a procedure using one of the most powerful techniques available for modelling species distributions, i.e. generalised additive modelling (e.g. Thuiller et al. 2003), to extrapolate original European plant, bird, mammal, reptile and amphibian species distribution atlas data from ~50x50 km grid to finer 10' grid resolution (i.e. ~10x10-km grid). The consequences of this downscaling procedure for

conservation planning are discussed with particular emphasis on its effects on: 1) spatial patterns of coincidence in species richness; 2) distribution of richness, rarity and complementarity hotspots.

DATA AND METHODS

Species data

Original species data included 868,960 records of occurrence for different groups of European terrestrial vertebrates and higher plants. These included 187 mammal (Mitchell-Jones et al. 1999), 445 breeding bird (Hagemeyer & Blair 1997), 149 amphibian and reptile (Gasc et al. 1997), and 2,362 plant species (Jalas & Suominen 1972-1996). Data vary with regard to taxonomic coverage. Terrestrial vertebrates include all known species, whereas plants comprise ~20% of the European flora (Humphries et al. 1999). The grid used is based on the Atlas Florae Europaeae (Lathi and Lampinen 1999), with cell boundaries typically following the 50 km lines of the Universal Transverse Mercator (UTM) grid, except near the border of the six-degree UTM zones and at coasts. The remaining atlases use slightly different grid-systems, including different rules to represent data on islands and coasts. Hence, vertebrate data were converted to the AFE grid system by identifying unique (though sometimes approximate) correspondence between cells in these grids. The mapped area (2,434 grid cells) includes western, northern and southern Europe, but excludes most of the eastern European countries (except for the Baltic States) where recording effort was both less uniform and less intensive (for more details on data conversion see (Williams et al. 2000)).

Climate and land-cover data

Climate and land cover data included 20 variables interpolated for Europe in the context of the EC funded ATEAM project (<http://www.pik-potsdam.de/ateam>). All data were developed at a spatial resolution of 10' for European grid cells based on the ATEAM geographic window and then aggregated to the Atlas Flora Europaeae 50x50 km grid. Climate data were averaged for the period of 1961-1990 and included mean annual temperature, mean temperature of the coldest month per year, mean annual precipitation sum, mean winter precipitation sum, mean summer precipitation sum, mean annual growing degree days, mean ratio of annual actual over annual potential evapotranspiration.

The land-cover data were derived by aggregation of the PELCOM land cover database. PELCOM is a 1-km pan-European land-cover database developed mainly from remotely sensed data. The classification methodology is based on a regional and integrated approach of the NOAA-AVHRR satellite data and ancillary information such as topographic features (Mücher 2000). Although finer spatial resolution databases exist such as CORINE (CEC 1993), PELCOM was selected because of its complete spatial coverage of the European window, and because of the

homogeneity of the methodology used for the land-cover classification. PELCOM is also the most up-to-date of pan-European land-cover databases. The percentages of each land-cover class were calculated for each individual 10' grid cells from the 1 km PELCOM cells. The data comprised four classes of land cover: forest, agriculture, urban and others. Forests were further subdivided into three classes: deciduous, coniferous and mixed forest. Agriculture was divided into two classes: the percentage of arable land and the percentage of grassland. The 'other' land cover class comprised: semi-natural areas (divided into areas of shrubland, and barren land), inland waters, wetlands, permanent ice and snow and the sea.

Downscaling process

Original species data were randomly divided into a calibration (70%) and an evaluation (30%) data set. Species with less than ten records in the calibration data set were excluded from analysis, so as to reduce errors associated with excessively small sample sizes (Stockwell & Peterson 2002). Generalised Additive Models (GAM, Hastie & Tibshirani 1990) were then used to extrapolate the original 50x50-km grid distributions to a 10' resolution. The modelling process was implemented in an iterative fashion, as described in Table 1. In this process, stepwise GAM models are run on the calibration data set using climate variables, while residuals are regressed with land cover. Selected land-cover variables for every species are then added to the set of selected climatic variables to produce alternative models including land cover. This was necessary because stepwise models including all variables at once would only rarely select land-cover variables (for a more detailed discussion see Thuiller et al. in press). The predictive accuracy of models using climate, or the combination of climate and land-cover variables, was assessed using the evaluation data set. Model accuracy was estimated with the area under the curve (AUC) index, based on the receiver-operating characteristic (ROC) curve (Fielding & Bell 1997). This index varies from 0.5 (random assignments) to 1 (accurate predictions) and represents the probability that the model correctly predicts the observed presences and absences. For each species, the model with greatest AUC was retained and used to extrapolate distributions to finer resolution. All analyses were performed with S-PLUS (Anon 1999).

Table 1

Assessing consequences of downscaling for reserve selection

The consequences of downscaling are assessed in terms of the models' ability to reproduce original observed patterns at coarser scale. Two kinds of comparisons are made: 1) Spearman rank correlations of species richness scores within and between groups; 2) analysis of the coincidence between hotspots for richness, rarity and complementarity before and after models are run.

Hotspots of richness are n top-ranking cells with highest sum values of species presence. Hotspots of rarity are n top-ranking cells with highest sum of species rarity values, measured as the sum of each species' inverse number of grid-cell records = $\sum (1/c_i)$, $i: c_i \neq 0, 1 \leq i \leq n$, where c_i is the number of grid cells occupied by species i (Araújo 1999; Williams et al. 2000). Hotspots of complementarity (Araújo & Williams 2001; Araújo et al. 2002a) are a combination of cells that maximise species representation for a given number of grid cells (also known as maximum coverage solution, Church et al. 1996). To calculate complementarity hotspots we used a heuristic technique adapted from the near-minimum-set algorithm of Margules et al. (1988). The algorithm starts by (1) selecting all areas with taxa that are equally or more restricted than the representation goal. For example, for a goal of representing each species at least once, it begins by selecting all areas that have species recorded in only one grid cell. Then (2) the algorithm follows a simple set of rules, applied iteratively to select areas richest in the rarest taxa. First it selects grid cells with the greatest complementary richness in just the rarest taxa (ignoring other taxa). If there are ties (3), it proceeds by selecting areas among ties that are richest in the next-rarest taxa. If there are still ties, it then selects those areas among ties with the lowest grid-cell number (this is an arbitrary rule used in place of random choice among ties in order to ensure repeatability in tests). Finally these steps are repeated as necessary until the representation goal is achieved. A test is performed to reject any grid cell that in hindsight is redundant to the selected goal. A modified version of this algorithm was used to provide an approximation to a maximum coverage problem. For this purpose, we repeated steps one to three until the required number of areas was attained or exceeded. A final re-ordering of areas by complementary richness was made to provide an approximate solution to the maximum-coverage problem (e.g. Araújo & Williams 2001; Williams et al. 2000). To ensure comparability between hotspots selected with downscaled and non-downscaled distributions a similar area was selected for both cases. These included 100 hotspots for non-downscaled distributions and 2500 hotspots for downscaled ones. Because there are usually a great number of equally efficient complementarity hotspots we selected 10 solutions for each taxonomic group by breaking ties randomly within each run. Hotspots were identified with WORLDMAP software (Williams 1999).

RESULTS

Generalised Additive Models provided generally good performances, i.e. with AUC values above 0.9 (Swets 1988) on the evaluation data set. Herptiles were the group showing the highest AUC values (Mean =0.95, SD =0.05), immediately followed by plants and birds (Mean =0.93, SD =0.04 and 0.05 respectively) and

then mammals (Mean =0.91, SD =0.05). The inclusion of land-cover variables increased the predictive performance of models 26% of the times. Downscaled distributions appeared to be able to recover fine spatial gradients (e.g. altitudinal and land cover) that were otherwise not apparent in the coarse 50 x 50 km grid scale (Figure 1).

Figure 1

By excluding species with less than 10 records in the calibration data, we eliminated 28% (954 species) of species in the data: 36% of plants (844 species), 12% of birds (55 species), 21% of mammals (40 species) and 10% of reptiles and amphibians (15 species). Exclusion of the rarest species in the original data did not affect overall patterns of richness within groups. That is, Spearman rank correlations between richness scores of data including all species and data excluding species for the models was 1 for birds, mammals and herptiles and was 0.99 for plants. Similarly, species richness correlations between groups were not greatly affected by exclusion of the rarest species, with cross-taxon correlations between plants (the group with the greatest proportion of excluded species) and the other groups being the weakest (Table 2a,b). Cross-taxon species richness correlations between downscaled group distributions shared the same broad patterns of coincidence in the original data, except for plants where a markedly different pattern emerged (Table 2b-c).

Table 2

The consequences of excluding rarest species were more serious for reserve selection. Here, only richness hotspots were not greatly affected by this exclusion, with hotspots selected with all data coinciding between 88 and 100% with hotspots selected with exclusion of the rarest species (Table 3). Coincidence was weaker for rarity hotspots and complementarity hotspots, where coincidences in hotspots ranged from 35-79% and 23-47% respectively.

Table 3

Although the total area reserved was comparable between downscaled and non-downscaled hotspots (i.e. 100 hotspots x ~50km x ~50km = 2500 hotspots x ~10km x ~10km) the latter solutions were able to span a wider range of areas in Europe (Figure 2-4). This is a logical consequence of using smaller but greater number of grid cells as selection units, since poor-biodiversity areas within large grid-cells are traded-off by more targeted areas in smaller grid cells. Nevertheless, non-downscaled and downscaled distributions selected hotspots that have a similar distribution in Europe. This is particularly true for most of the hotspots selected with non-downscaled species data, which tend to be included by downscaled

hotspots. There are, however, some displacements in the location of hotspots selected with all the three methods. Examples include, for instance, complementarity hotspots for mammals that select priority areas in southern-western Finland with non-downscaled distributions, but not with downscaled data that selected hotspots on the eastern fringes of Norway, instead. Similarly, richness hotspots among non-downscaled bird distributions include areas in Bulgaria and Romania, while downscaled hotspots of richness excluded any area near such lower latitudes.

Another difference between non-downscaled and downscaled hotspot distribution is that at higher resolution the patterns of species overlap are seen more precisely to be greater at mid elevations. For example, richness hotspots for mammals and plants have a tendency to concentrate on central European mountains like the Pyrenees, Alps, and Carpathians for both downscaled and non-downscaled solutions. However, non-downscaled solutions appeared to select richness hotspots on mountaintops (an artefact caused by the large grid-cell size), while downscaled solutions seem to favour mid-elevation sections of these mountains.

Figure 2-4

DISCUSSION

By downscaling species atlas distributions to a finer resolution there are gains and losses. The first loss is the exclusion of restricted-range species from analysis (but see He & Gaston 2000; Kunin 1998). This is required because the kind of interpolation techniques used here are unable to model species with only a few records (Stockwell & Peterson 2002). The second loss is related to the uncertainty of model predictions, which may accumulate error through the process of overlaying modelled distributions for many species (e.g. Flather et al. 1997). It is generally accepted that original species atlas distribution data provide reliable estimates of presence (false positives are potentially low), although the same cannot be said about absences (false negatives are potentially high) (e.g. Araújo & Williams 2000). Model predictions add a new degree of uncertainty given that the number of false positives is likely to increase to a variable but unknown extent, while the number of false negatives, although expected to diminish, remains unknown. This problem is particularly severe when models are used to extrapolate distributions into a different resolution (the case of this paper), regions (e.g. Peterson & Vieglais 2001), or times (e.g. Thuiller et al. in press), because model evaluation techniques are limited to the original circumstances upon which the models were built (Beutel et al. 1999).

However, distribution atlases are often too coarse to be used in most conservation-planning applications and downscaling procedures remain attractive despite uncertainties (e.g. Barbosa et al. 2003; Boitani et al.

1999; Wisz, in preparation). The argument is often made that many urgent problems in conservation biology require that decisions be made with available data and models so that "... the choice is between crudely imperfect advice and none at all" (May 1984 cited by Flather et al. 1997). The opposing and perhaps not very useful argument is that one should not assume that crude advice is better than no advice (Slobodkin 1988). In practice, there is a possibility that downscaled procedures might recover fine-grain species-distribution patterns with an acceptable accuracy. This being true they would help locating conservation areas that are more precisely related to environmental features of the land, hence providing an incentive for using existing datasets in real-world conservation planning. This appeared to be the case of our particular application where downscaled maps of species richness were able to recover major topographical (e.g. altitudinal gradients) and climate gradients of species richness otherwise lost from original coarse-scale resolution atlases. Between-group downscaled patterns of terrestrial vertebrate species richness were also broadly coincident between downscaled models and original coarse-resolution data (except for plants), which suggest that downscaling did not change the original structure in the data for these groups.

Given the opportunities and uncertainties arising from downscaling procedures, the usefulness of the process is likely to be contingent on the goals and the ability to test results with targeted expert-judgements or field-testing of the predictions. Qualitative or quantitative assessments of results would be particularly important when downscaled distributions are used to support local conservation planning decisions, because there is a risk that downscaled maps might yield spurious predictions of where species occur. The critical issue is whether errors arising from accepting false positives (i.e. model predictions where species do not occur) are more severe than errors arising from false negatives (i.e. coarse-resolution atlases with incomplete representations of species distributions), or from ignorance of the local heterogeneity of species distributions within coarser-resolution grid cells. There are no established rules to find the appropriate trade-off between such opposing sources error. Nevertheless it seems reasonable to expect that areas with very poor datasets might benefit from using modelled data (whether it is downscaled or not) whereas data-rich areas might have more to lose with adding model uncertainties in the decision process.

ACKNOWLEDGEMENTS

Species distribution data was kindly supplied by J.P. Gasc (herptiles), W.J.M. Hagemeyer (breeding birds), Raino Lampinen (plants); and A.J. Mitchell-Jones (mammals). We thank Tim Mitchell for generating the climate data; Markus Erhard and Sönke Zaehle for aggregating it into the AFE grid. This research is

funded by EC ATEAM and ALARM projects; MBA also thanks FCT (SFRH/BPD/5547/2001) for funding.

REFERENCES

- Anon 1999. S-PLUS 2000 User's Guide. MathSoft, Seattle, WA.
- Araújo, M. B. 1999. Distribution patterns of biodiversity and the design of a representative reserve network in Portugal. *Diversity and Distributions* **5**:151-163.
- Araújo, M. B. 2003. The coincidence of people and biodiversity in Europe. *Global Ecology and Biogeography* **12**:5-12.
- Araújo, M. B., C. J. Humphries, P. J. Densham, R. Lampinen, W. J. M. Hagemeyer, A. J. Mitchell-Jones, and J. P. Gasc. 2001. Would environmental diversity be a good surrogate for species diversity? *Ecography* **24**:103-110.
- Araújo, M. B., and P. Williams. 2000. Selecting areas for species persistence using occurrence data. *Biological Conservation* **96**:331-345.
- Araújo, M. B., and P. H. Williams. 2001. The bias of complementarity hotspots toward marginal populations. *Conservation Biology* **15**:1710-1720.
- Araújo, M. B., P. H. Williams, and A. Turner. 2002a. A sequential approach to minimise threats within selected conservation areas. *Biodiversity and Conservation* **11**:1011-1024.
- Araújo, M. B., S. E. Williams, and R. J. Fuller. 2002b. Dynamics of extinction and the selection of nature reserves. *Proceedings of the Royal Society London B* **269**:1971-1980.
- Barbosa, A. M., R. Real, J. Olivero, and J. M. Vargas. 2003. Otter (*Lutra lutra*) distribution modelling at two resolution scales suited to conservation planning in the Iberian Peninsula. *Biological Conservation* **114**:377-387.
- Beutel, T. S., R. J. S. Beeton, and G. S. Baxter. 1999. Building better wildlife - habitat models. *Ecography* **22**:219-223.
- Boitani, L., F. Corsi, I. De Biase, D. I. C., M. Ravagli, G. Reggiani, I. Sinibaldi, and P. Trapanese. 1999. A databank for the conservation and management of the African Mammals. Instituto Ecologia Applicata, Roma.
- CEC 1993. CORINE Land Cover technical guide. European Union. Directorate-General Environment, Nuclear Safety and Civil Protection., Luxemburg: Office for Official Publications of the European Communities.
- Church, R. L., D. M. Stoms, and F. W. Davis. 1996. Reserve selection as a maximal covering location problem. *Biological Conservation* **76**:105-112.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**:38-49.
- Flather, C. H., K. R. Wilson, D. J. Dean, and W. C. McComb. 1997. Identifying gaps in conservation networks: of indicators and uncertainty in geographic-based analysis. *Ecological Applications* **7**:531-542.
- Freitag, S., A. O. Nicholls, and A. S. van Jaarseveld. 1998. Dealing with established reserve networks and incomplete distribution data sets in conservation planning. *South African Journal of Science* **94**:79-86.
- Gasc, J.-P., A. Cabela, J. Crnobrnja-Isailovic, D. Dolmen, K. Grossenbacher, P. Haffner, J. Lescure, H. Martens, J. P. Martinez Rica, H. Maurin, M. E. Oliveira, T. S. Sofianidou, M. Veith, and A. Zuiderwijk. 1997. Atlas of amphibians and reptiles in Europe. Societas Europaea Herpetologica & Museum National d'Histoire Naturelle, Paris.
- Hagemeyer, W. J. M., and M. J. Blair. 1997. The EBCC atlas of European breeding birds, their distribution and abundance. Poyser, London.
- Hastie, T. J., and R. Tibshirani. 1990. Generalized additive models. Chapman and Hall, London.
- He, F., and K. Gaston. 2000. Estimating species abundance from occurrence. *The American Naturalist* **156**:553-559.
- Hopkinson, P., J. Evans, and R. D. Gregory. 2000. National-scale conservation assessment at an appropriate resolution. *Diversity and Distributions* **6**:195-204.
- Humphries, C., M. B. Araújo, P. Williams, R. Lampinen, T. Lahti, and P. Uotila. 1999. Plant diversity in Europe: *Atlas Florae Europaeae* and *WORLDMAP*. *Acta Botanica Fennica* **162**:11-21.
- Jalas, J., and J. Suominen. 1972-1996. *Atlas Florae Europaeae*. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.
- Koleff, P., and K. J. Gaston. 2002. The relationship between local and regional species richness and spatial turnover. *Global Ecology & Biogeography* **11**:363-375.
- Kunin, W. E. 1998. Extrapolating species abundance across spatial scales. *Science* **281**:1513-1515.
- Margules, C. R., A. O. Nicholls, and R. L. Pressey. 1988. Selecting networks of reserves to maximise biological diversity. *Biological Conservation* **43**:63-76.
- Mitchell-Jones, A. J., G. Amori, W. Bogdanowicz, B. Krystufek, P. J. H. Reijnders, F. Spitzenberger, M. Stubbe, J. M. B. Thissen, V. Vohralik, and J. Zima. 1999. Atlas of European Mammals. Academic Press, London.
- Mücher, C. A., editor. 2000. PELCOM project 'development of a consistent methodology to derive land cover information on a European scale from remote sensing for environmental modelling'. Final report Febr. 2000, part 2, EU DG XII-D.
- Peterson, A. T., and D. A. Vieglais. 2001. Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *BioScience* **51**:363-371.

- Rebelo, A. G., and W. R. Siegfried. 1992. Where should nature reserves be located in the Cape Floristic Region, South Africa? Models for the spatial configuration of a reserve network aimed at maximising the protection of floral diversity. *Conservation Biology* **6**:243-252.
- Rouget, M. 2003. Measuring conservation value at fine and broad scales: implications for a diverse and fragmented region, the Agulhas Plain. *Biological Conservation* **112**:217-232.
- Slobodkin, L. B. 1988. Intellectual problems of applied ecology. *BioScience* **38**:337-342.
- Stockwell, D. R. B., and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* **148**:1-13.
- Stoms, D. M. 1992. Effects of habitat map generalization in biodiversity assessment. *Photogrammetric Engineering & Remote Sensing* **58**.
- Stoms, D. M. 1994. Scale dependence of species richness maps. *Professional Geographer* **46**:346-358.
- Swets, K. A. 1988. Measuring the accuracy of diagnostic systems. *Science* **240**:1285-1293.
- Thuiller, W., M. B. Araújo, and S. Lavorel. 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* **14**:669-680.
- Thuiller, W., M. B. Araújo, and S. Lavorel. in press. Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography*.
- Thuiller, W., L. Brotons, M. B. Araújo, and S. Lavorel. In press. Effects of restricting environmental range of data to project current and future species distributions. *Ecography*.
- Turner, J. R. G., J. J. Lennon, and J. A. Lawrenson. 1988. British bird species distributions and the energy theory. *Nature* **335**:539-541.
- van Rensburg, B. J., S. L. Chown, and K. J. Gaston. 2002. Species richness, environmental correlates, and spatial scale: a test using South African Birds. *The American Naturalist* **159**:566-577.
- Williams, P., C. Humphries, M. B. Araújo, R. Lampinen, W. Hagemeyer, J.-P. Gasc, and T. Mitchell-Jones. 2000. Endemism and important areas for representing European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belgian Journal of entomology* **2**:21-46.
- Williams, P. H. 1999. *WORLDMAP 4 Windows: Software and help document 4.2*. Distributed privately, London.
- Williams, P. H., H. M. d. Klerk, and T. M. Crowe. 1999. Interpreting biogeographical boundaries among Afrotropical birds: spatial patterns in richness gradients and species replacement. *Journal of Biogeography* **26**:459-474.

Table 1 – Iterative downscaling process

Step	Rule or procedure
1	Run stepwise GAM using climate variables only
2	Run stepwise GAM on the residuals of step 1 using land cover as predictor variables
3	Run GAM using selected climatic and land cover variables in steps 1 and 2
4	Compare predictive accuracies in models of steps 1 and 3 with ROC
5	For each species retain the model with better accuracy
6	Use retained models to extrapolate distributions at 10 ⁷ resolution.

Table 2 – Spearman rank correlation coefficients of species richness scores among groups of plants and terrestrial vertebrates in Europe with (a) original data including all species; (b) with modelled data excluding species with less than 10 records in the calibration set; c) downscaled data excluding species with less than 10 records in the calibration set

	Plants	Birds	Mammals	Herptiles
(a)				
Plants	1			
Birds	0.46	1		
Mammals	0.75	0.63	1	
Herptiles	0.49	0.08	0.41	1
b)				
Plants	1			
Birds	0.49	1		
Mammals	0.76	0.63	1	
Herptiles	0.42	0.09	0.41	1
(c)				
Plants	1			
Birds	0.06	1		
Mammals	0.64	0.61	1	
Herptiles	0.78	-0.09	0.44	1

Table 3 – Coincidence between hotspots selected with all species and with modelled species alone. For richness and rarity hotspots numbers are percentages of coincident hotspots. For complementarity hotspots numbers are mean percentages of coincident hotspots, after ten simulations, while numbers within brackets are standard deviations to the mean.

	Richness hotspots	Rarity hotspots	Complementarity hotspots
Plants	88	79	47 (1.7)
Breeding birds	100	53	23 (2.3)
Mammals	100	35	29 (4.3)
Herptiles	96	49	29 (4.1)

FIGURE LEGENDS

Figure 1 – Species richness scores among downscaled and non-downscaled distributions for a) plants; b) breeding birds; c) mammals; and d) herptiles. Species with less than 10 records in the calibration data sets for models are not plotted on these maps. We used a 6 grey-scale class, where increasing intensities of brown represent increasing richness scores and increasing intensities of light brown represent decreasing richness scores.

Figure 2 – Distribution of downscaled and non-downscaled hotspots rarity in Europe

Figure 3 – Distribution of downscaled and non-downscaled hotspots richness in Europe

Figure 4 – Distribution of downscaled and non-downscaled hotspots of complementarity in Europe

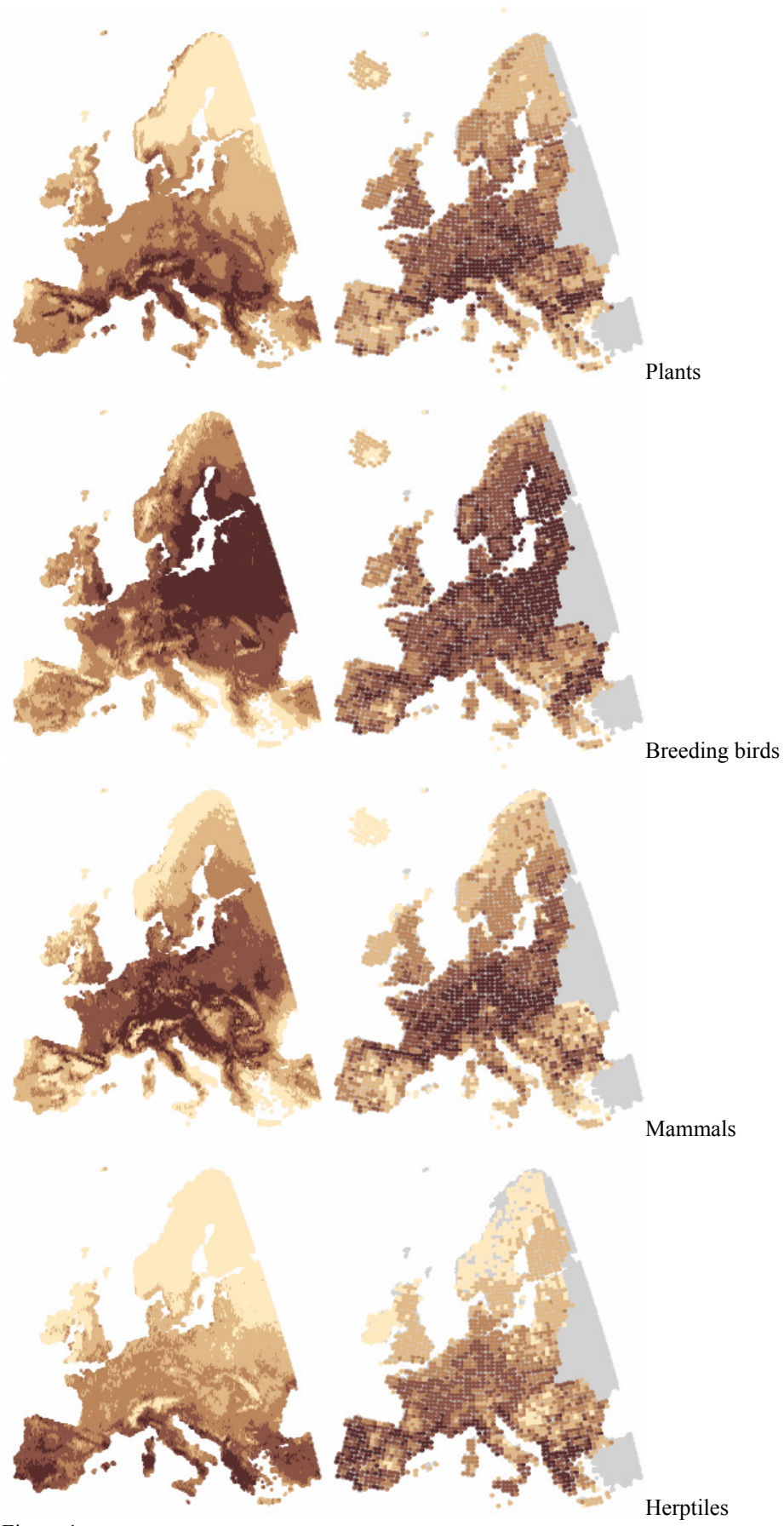


Figure 1 –

Downscaled hotspots

Non-Downscaled hotspots

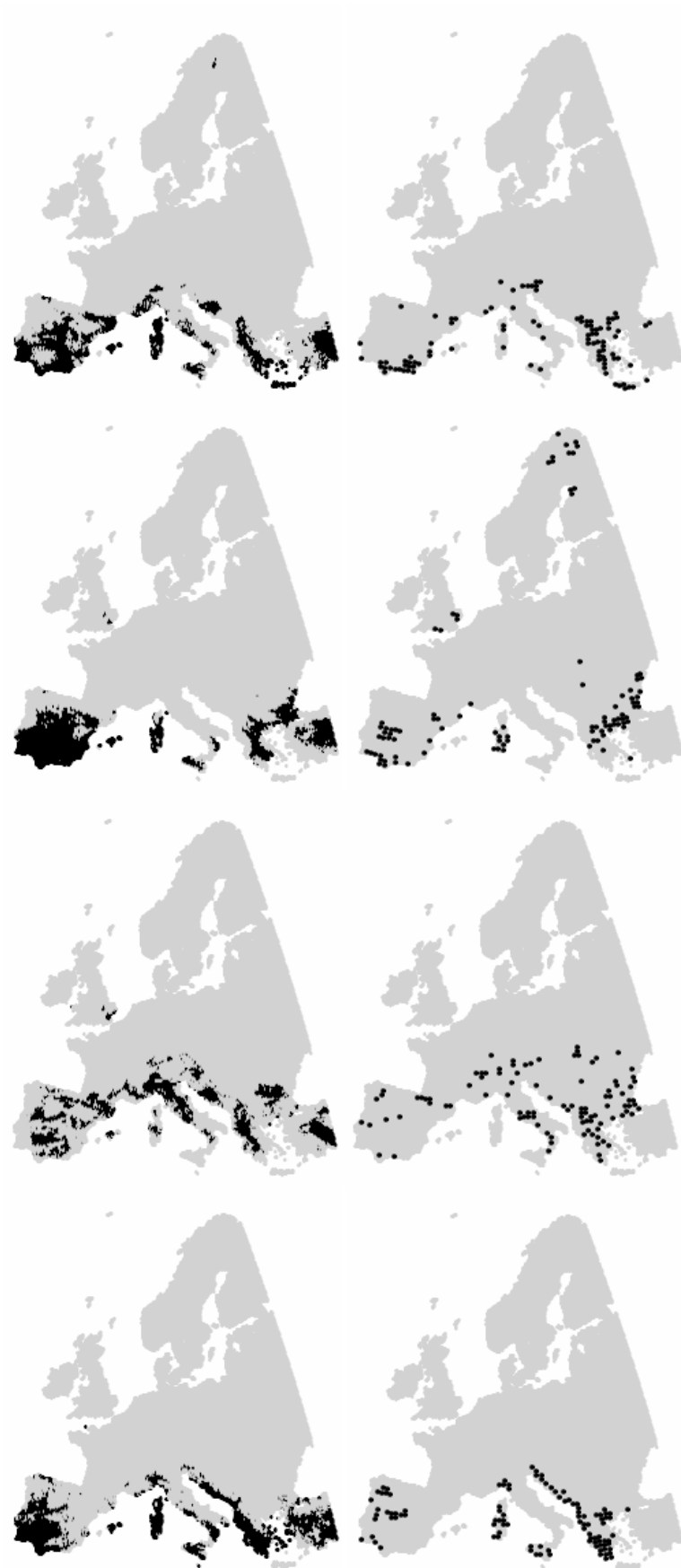


Figure 2 –

Downscaled hotspots

Non-Downscaled hotspots

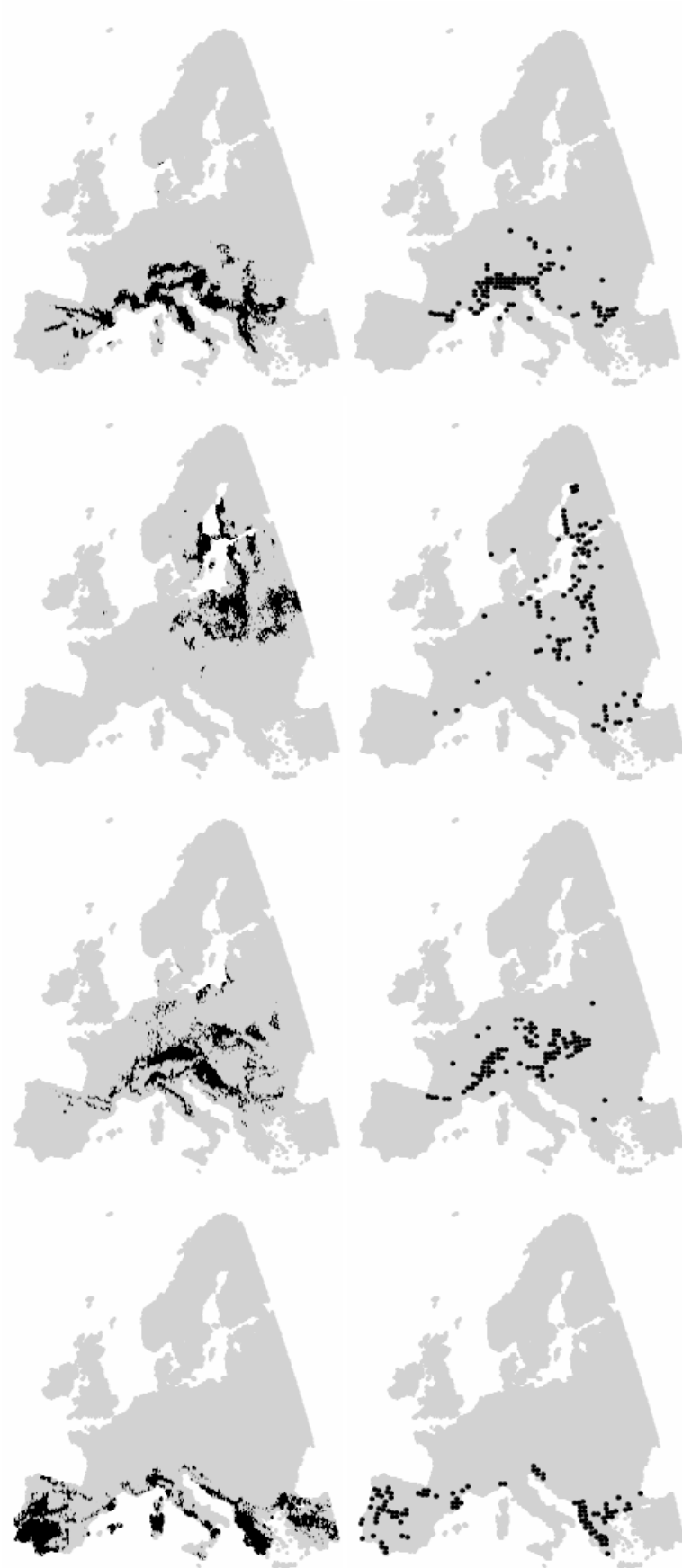


Figure 3 –

Downscaled hotspots

Non-Downscaled hotspots



Figure 4 –

ANNEXE 2.1

BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change

Thuiller W.

Global Change Biology (2003) **9**, 1353-1362

BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change

WILFRIED THUILLER

Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de Recherche Scientifique, 1919 route de Mende, 34293 Montpellier Cedex 5, France

Abstract

A new computation framework (BIOMOD: BIOdiversity MODelling) is presented, which aims to maximize the predictive accuracy of current species distributions and the reliability of future potential distributions using different types of statistical modelling methods. BIOMOD capitalizes on the different techniques used in static modelling to provide spatial predictions. It computes, for each species and in the same package, the four most widely used modelling techniques in species predictions, namely Generalized Linear Models (GLM), Generalized Additive Models (GAM), Classification and Regression Tree analysis (CART) and Artificial Neural Networks (ANN). BIOMOD was applied to 61 species of trees in Europe using climatic quantities as explanatory variables of current distributions. On average, all the different modelling methods yielded very good agreement between observed and predicted distributions. However, the relative performance of different techniques was idiosyncratic across species, suggesting that the most accurate model varies between species. The results of this evaluation also highlight that slight differences between current predictions from different modelling techniques are exacerbated in future projections. Therefore, it is difficult to assess the reliability of alternative projections without validation techniques or expert opinion. It is concluded that rather than using a single modelling technique to predict the distribution of several species, it would be more reliable to use a framework assessing different models for each species and selecting the most accurate one using both evaluation methods and expert knowledge.

Keywords: artificial neural network, BIOMOD, climate change, future climatic envelopes, generalized additive model, generalized linear model, regression tree analysis

Received 2 April 2003; revised version received and accepted 3 May 2003

Introduction

Recently documented biotic responses to possible climate and land-use change (Walther *et al.*, 2002, Peterson *et al.*, 2002b; Midgley *et al.*, 2003) raise a crucial question for ecologists and conservationists: Are projected climate and land-use change likely to threaten biodiversity and conservation of species? Several meta-analyses have demonstrated that there is a globally coherent 'fingerprint' of climate change impacts across natural systems and particularly on biodiversity (Parmesan & Yohe, 2003; Root *et al.*, 2003). Static

modelling techniques have been used to assess the impacts of global change on biodiversity distribution, by predicting current species distributions and applying statistical models from current distributions to project future distributions under global change scenarios (for a review, see Pearson & Dawson, 2003). Static modelling relates the current observed distribution of species to a pool of available environmental variables using statistical or rule-based models. Different modelling techniques are commonly applied to project potential future species distributions. Although Generalized Linear Models (GLM) are the most common (Hill *et al.*, 1999; Bakkenes *et al.*, 2002), Generalized Additive Models (GAM) are being used increasingly (Leathwick, 1995; Frescino *et al.*, 2001).

Correspondence: Wilfried Thuiller,
e-mail: wilfried.thuiller@cefe.cnrs-mop.fr

Alternative rule-based approaches have also emerged as interesting tools to predict current and future potential distributions. These include Classification and Regression Tree analysis (CART) (Rouget *et al.*, 2001; Thuiller *et al.*, 2003b) or Artificial Neural Networks (ANN) (Pearson *et al.*, 2002; Moisen & Frescino, 2002). Modellers have observed that different modelling techniques for the same species may give different results and that different methods tend to vary idiosyncratically across species (Thuiller *et al.*, 2003a). While some comparative studies may guide methodological choices (e.g. Manel *et al.*, 1999; Thuiller *et al.*, 2003a), it may be obscure for non-statisticians or inexperienced users to select a modelling technique to predict species distributions for a specific application.

The purpose of this paper is therefore to present a modelling application, BIOMOD, that allows spatial predictions maximizing the model accuracy and projected future species distributions. Indeed, to predict current species distributions and to project them into the future, it may be necessary to consider more than one modelling technique and use the most accurate one for each species rather than the most accurate across species. In addition, models with similar accuracies for current distributions could have very different behaviour when applying them to future environmental parameter sets. The general framework that is proposed uses different modelling techniques from parametric statistical to non-parametric rule-based (GLM, GAM, CART and ANN) for each species, to predict current species distributions, to evaluate them and then either choose only the most accurate model to make future projections for each species, or keep all the predictions from the different models and project all of them into the future.

Methods

Species data

Sixty-one native tree taxa distributed across Europe were considered for modelling. This covers most of the important timber taxa of Europe, including all the gymnosperm softwoods (Pinales, Taxales and Gnetales) and some hardwoods (Myricales, Juglandales and Fagales) (Humphries *et al.*, 1999). Trees were chosen because: (i) their distribution and ecology is relatively well known compared with other plant taxa; (ii) their richness is correlated (Spearman correlation $\rho = 0.80$, $P < 0.001$) with the overall richness of the Atlas Flora Europaeae (AFE) data set (Araújo & Williams, 2000); and (iii) they are long-lived organisms and their distribution is relatively stable in comparison with

some other groups. The species presence–absence data are a subset of AFE (Jalas & Suominen, 1972–1996), which was digitized by Lahti & Lampinen (1999). Data are located in 4419 UTM (Universal Transverse Mercator) 50×50 -km² grid cells. We used only 2089 grid cells, excluding most of the eastern European countries (except for the Baltic States) because of low recording efforts in these areas (Williams *et al.*, 2000).

Environmental data

The climate data used for analyses are a comprehensive set of bioclimatic variables for Europe from the Climatic Research Unit (<http://www.cru.uea.ac.uk/>) (Mitchell, 2002) (<http://www.pik-potsdam.de/ateam/>): mean annual, winter and summer precipitation, mean annual temperature and minimum temperature of the coldest month, growing degree days ($>5^\circ$) and an index of humidity (mean ratio of annual actual over annual potential evapotranspiration). Means are averaged for the period of 1961–1990. These data sets are supplied on a 10' grid, covering Europe. Then data were aggregated by averaging to 50×50 -km² UTM in order to match the resolution of species data.

To examine future species distributions, I used the climate data obtained in a 10' resolution for 2050 based on the GCM experiments conducted at the UK Hadley Centre for Climate Prediction and Research using the HadCM3 model (Mitchell, 2002) under the SRES scenario A1FI (Nakicenovic & Swart, 2000).

BIOMOD – model calibrations

In order to evaluate the quality of predictions, we divided databases into two subsets: calibration and evaluation. The first, a random sample from 70% of the total database, was used to calibrate (train) the models, whereas the second, comprising the remaining data, was used to evaluate (test) model predictions (Fielding & Bell, 1997). The methods included in the testing framework were GLM, GAM, CART and ANN. GLM are the most commonly used technique in species distribution modelling and have shown good ability to predict current species distributions (Austin & Meyers, 1996; Brito *et al.*, 1999). However, the inability of GLM to deal with complex response curves (Yee & Mitchell, 1991) has stimulated the use of GAM in ecological modelling (Lehmann *et al.*, 2003; Thuiller *et al.*, 2003a). CART are less commonly used than the previous two methods, but are accurate and useful to describe hierarchical interactions between species (Franklin, 1998; Rouget *et al.*, 2001; Thuiller *et al.*, 2003). ANN are being increasingly used (Pearson *et al.*, 2002), but they are still limited by the difficulty in identifying

causal relationships and dominant input variables from the network structure. However, their ability to deal with non-normal statistical distributions and their adequacy to determine environmental envelopes that have non-linear responses to environmental variables offer considerable advantages.

For each species, the following procedures were used:

- (1) GLM (McCullagh & Nelder, 1989) with linear, second and third order and polynomial terms (second and third order) with the possibility of interactions only for the linear terms. A stepwise procedure was used to select the most significant variables using the AIC criterion (Akaike, 1974; Anon, 1999).
- (2) GAM (Hastie & Tibshirani, 1990) with smooth splines. The degree of smoothness was automatically selected by cross-validation and bounded to 4 for each variable. As for GLM, a stepwise procedure was used to select the most parsimonious model using the AIC criterion (Akaike, 1974; Anon, 1999). GAM being, by definition, additive, interaction terms were not included.
- (3) CART (Breiman *et al.*, 1984) using the *rpart* library of Splus (Therneau & Atkinson, 1997). This procedure runs a 10-fold cross-validation to select the best trade-off between the number of leaves of the tree and the explained deviance.
- (4) ANN using the library *nnet* of Splus (Venables & Ripley, 2002). This ANN is a feed-forward neural network. ANN was parameterized using seven hidden units in a single hidden layer (selected by cross-validation), with a weight decay equal to 0.03. As each simulation gave slightly different results, the ANN was run 10 times and the mean was used to provide predictions and projections (Ripley, personal communication).

BIOMOD – model evaluations

The accuracy of models was evaluated using two different methods according to their specificity between ‘liberal’ and ‘conservative’ approaches, which have rather different implications in practice (Thuiller *et al.*, in review-b). First, the area under the relative operating characteristic curve (AUC) was computed (for more details, see Thuiller *et al.*, 2003a). This does not require the calculation of a threshold to transform probability values from models to binary presence–absence form (Pearce & Ferrier, 2000; Thuiller *et al.*, in review-b). Second, the κ statistics (Cohen, 1960) was calculated using a threshold optimizing this statistic (Monserud & Leemans, 1992; Thuiller *et al.*, in review-b).

BIOMOD – model predictions and projections

Each model was run for every species, compared and used to derive spatial predictions on the original data (2089 grid cells). Then calibrated models were used to project species distributions using the set of climatic variables for entire Europe at $10' \times 10'$. Potential future distributions were also projected by changing climate as predicted by the HadCM3 GCM under one SRES scenario (A1FI).

To transform probability values from each model, either of the following were used: the threshold maximizing the κ statistics, or the threshold maximizing simultaneously the number of presences and absences correctly predicted (Thuiller *et al.*, in review-b).

Results

Model evaluations

A test of the models was provided by evaluating their accuracy in predicting observed species distributions (Table 1). According to Monserud & Leemans (1992)

Table 1 Mean and standard deviation of AUC and κ statistics for the 61 species for each model, according to the data used

	AUC									κ								
	Cal			Eval			Pred			Cal			Eval			Pred		
	min	me	max	min	me	max	min	me	max	min	me	max	min	me	max	min	me	max
GLM	0.82	0.95	0.99	0.82	0.94	1.00	0.82	0.95	0.99	0.39	0.69	0.88	0.36	0.66	0.86	0.38	0.68	0.87
GAM	0.84	0.96	0.99	0.84	0.94	1.00	0.84	0.95	0.99	0.35	0.70	0.88	0.40	0.66	0.95	0.38	0.68	0.88
CART	0.83	0.94	0.99	0.54	0.87	0.96	0.82	0.92	0.98	0.45	0.72	0.88	0.27	0.59	0.84	0.45	0.68	0.86
ANN	0.87	0.97	1.00	0.82	0.95	0.99	0.85	0.96	0.99	0.54	0.79	0.94	0.34	0.68	0.84	0.50	0.75	0.89

Cal: calibration data; Eval: evaluation data; Pred: original data (calibration + evaluation). min, me and max are, respectively, minimum, mean and maximum values of AUC or κ statistics.

subjective guidelines, κ values above 0.7 can be considered to indicate very good agreement. On average, predicted distributions on both calibration and evaluation data were thus found to exhibit high levels of agreement with observed distributions. Conclusions are similar according to AUC, where subjective guidelines (Swets, 1988) suggest a very good agreement for AUC above 0.9. According to the guidelines presented for the two methods, ANN appeared slightly better than the other methods, while classification trees appeared as the weakest method in terms of discrimi-

Table 2 Percentage of best models across the 61 tree species according to both AUC and κ statistics carried out on evaluation data

	AUC	κ
GLM	14.75	21.31
GAM	24.59	31.15
CART	00.00	00.00
ANN	60.66	47.54

nation ability. Comparing the results for calibration, evaluation and the entire data, the tendency for CART to over-fit during the calibration process resulted in a generally poor agreement for evaluation data (Table 1). ANN also tended to over-fit during the calibration process, but the predictions on evaluation data still showed good agreement. On the other hand, GAM or GLM displayed very small differences in fit between calibration and evaluation data, and thus did not appear to over-fit.

Even if ANN exhibited a higher mean accuracy according to both the ROC curve and κ statistics, there were noticeable differences across species (Table 2). GLM or GAM had a higher accuracy than ANN in several cases, supporting the idea that there is no universal 'best' modelling technique.

Model predictions and projections

Predictions for the 61 modelled species revealed responses that were highly species- and modelling technique-dependent. Generally, the expected trend of

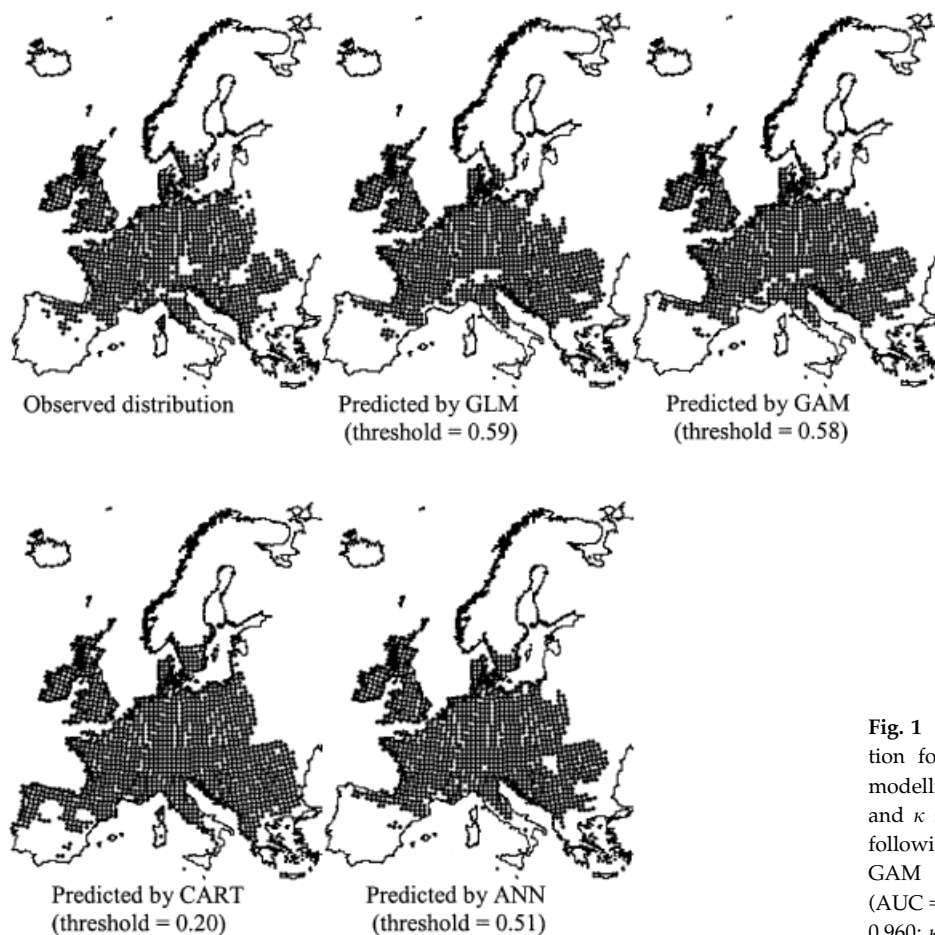


Fig. 1 Observed and predicted distribution for *Quercus petraea* according to the modelling technique used. The ROC curve and κ statistics on evaluation data are the following: GLM (AUC = 0.939; κ = 0.74), GAM (AUC = 0.945; κ = 0.77), CART (AUC = 0.927, κ = 0.76) and ANN (AUC = 0.960; κ = 0.79).

over-fitting by CART was quite obvious, and for a number of species GLM, GAM and ANN differed noticeably. Concerning the projections into the future, the expected trend of northward expansion was apparent for a number of species, but especially for northern native species (e.g. *Betula nana*). However, for many species, changes in potential distribution tended to be multidirectional. To illustrate the study, four species were selected to represent different chorological status in Europe: a Euro-Siberian species (*Quercus petraea*), a continental species (*Castanea sativa*), a Mediterranean species (*Pinus halepensis*) and a Siberian species (*B. nana*).

Predictions of current distribution for *Q. petraea* showed interesting differences according to the modelling technique (Fig. 1). First, even if both AUC and κ yielded similar results across models, they resulted in non-negligible differences in terms of spatial predictions. This indicates that small differences according to the evaluating statistic might translate into significant differences for spatial predictions.

Projections into the future highlighted that small differences for predicting current distributions are exacerbated in the future (Fig. 2). Future distribution projected by CART showed a northward expansion of *Q. petraea* similar to the other three modelling techniques. However, CART predicted a stable distribution in Spain and southern France while the other three methods showed a decrease of suitable conditions in these regions. Even if GLM, GAM and ANN showed similar trends, there were differences in the UK, northern France and Poland. According to the ranking provided by AUC and κ , the future potential distribution of *Q. petraea* projected by ANN should be the most accurate.

Differences between future potential species distributions according to the modelling technique used had a crucial impact to assess the sensitivity of a given species to climate change. For instance, the four models did not result in the same spatial projections in terms of suitable sites lost or gained to the future (Figs 3–5). The impact of climate change for *C. sativa* was different

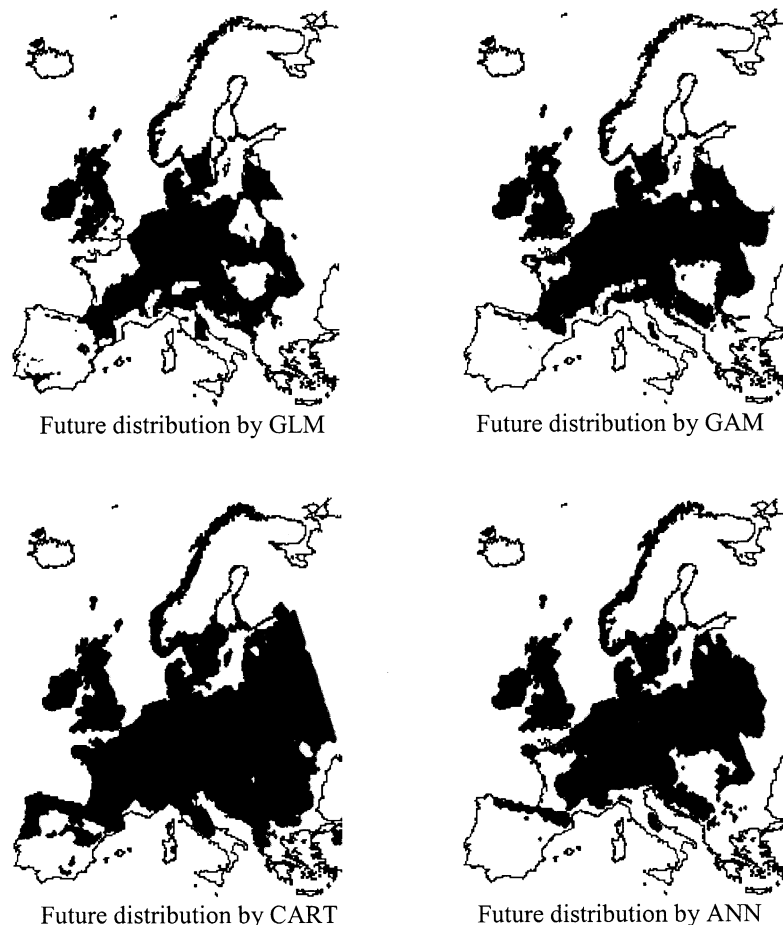


Fig. 2 Projected potential future distribution for *Quercus petraea* according to the modelling techniques. The thresholds used to transform probability values into binary presence-absence form are the same as for Fig. 1.

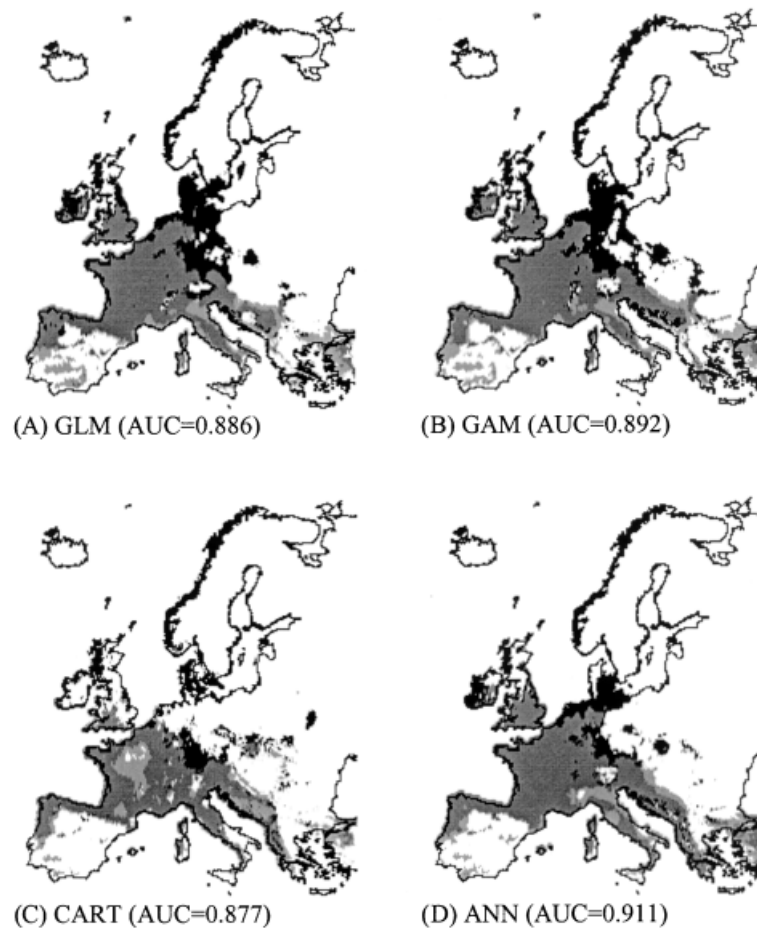


Fig. 3 Comparison of current projected distribution of *Castanea sativa* and its future potential distribution in 2050. Dark grey colour corresponds to stable suitable sites, grey to loss of suitable sites, black to gain of suitable sites and clear grey to stable unsuitable sites.

according to the modelling technique (Fig. 3). CART projected a stable distribution of *C. sativa*, while ANN, GAM and GLM projected a northward expansion and a contraction from southern Europe. For this species, according to the evaluating methods, ANN appeared to be the most accurate model to predict the current distribution and would therefore be the most reliable to generate future projections. Results for *P. halepensis*, a Mediterranean tree, also exhibit substantial discrepancies according to the models used (Fig. 4). For this species, there was a large difference between statistical and rule-based methods. GLM and GAM projected a slight northward expansion with contraction from the extreme southeast and southwest of its current range, while CART and ANN projected only a strong northward expansion without southern contraction. According to the evaluating methods, projection by GAM was the most reliable.

When modelling techniques yield very similar results in terms of accuracy, the results of sensitivity of species

to climate change show similar patterns. This was the case, for example, for *B. nana*, a Euro-Siberian species that was projected to lose a considerable amount of suitable habitat in the southern part of its distribution almost identically by all modelling techniques (Fig. 5).

Discussion

Results presented in this study suggest that the reliability of future potential species distributions depends strongly on the modelling technique used. A small difference between two modelling techniques in terms of AUC or κ can result in a large difference in projections of future potential habitat. It is worth noting that only a subsample of possible forms for GLM, GAM, CTA and ANN was used. The way in which the different modelling techniques in BIOMOD were parameterized is explained and detailed here and in other papers (Thuiller *et al.*, 2003a,b). However, there are different methods to parameterize these models; for

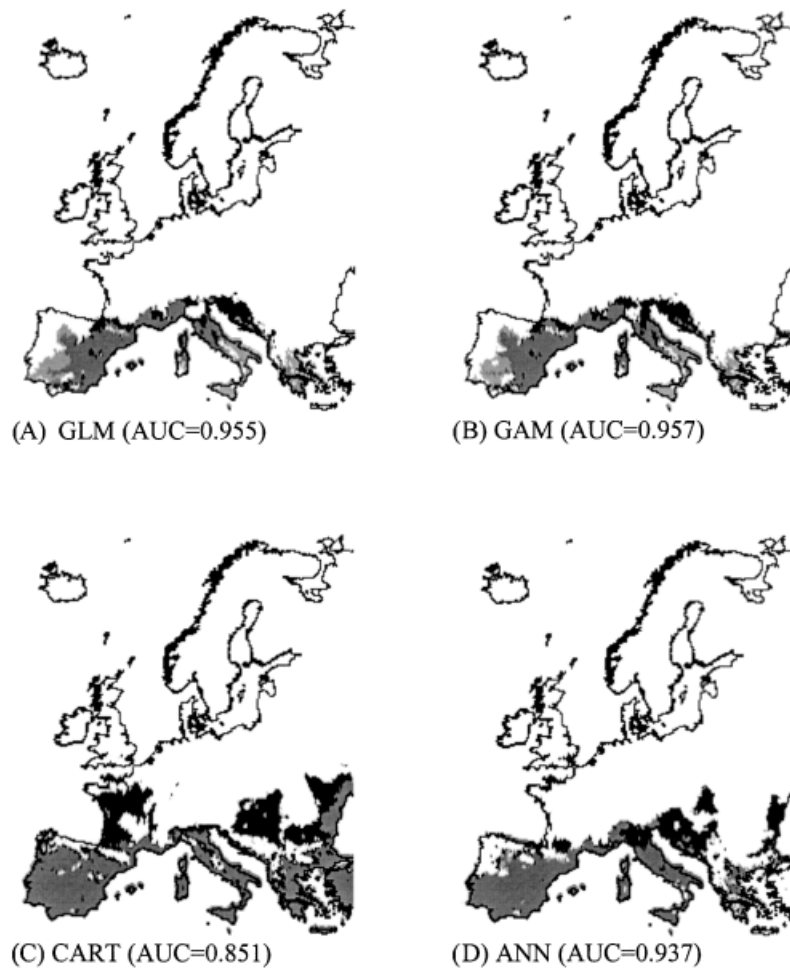


Fig. 4 Comparison of current projected distribution of *Pinus halepensis* and its future potential distribution in 2050. Dark grey colour corresponds to stable suitable sites, grey to loss of suitable sites, black to gain of suitable sites and clear grey to stable unsuitable sites.

instance, using lowess smoother for GAM instead of spline, using only linear terms in GLM, or using a higher number of hidden network units in ANN. A sensitivity analysis would be worthy to assess the importance of such parameters, but this was clearly beyond the scope of this paper. Moreover, if differences also exist according to the methods to fit a single model, it enhances the use of a framework such as BIOMOD that tests different models and chooses the most accurate one. BIOMOD could be used to test the same modelling technique, but with different forms and parameterizations. Readers interested in more detailed discussions and model comparisons could refer to Brotons *et al.* (in review), Franklin (1998), Guisan *et al.* (1999), Vayssières *et al.* (2000) and Thuiller *et al.* (2003a). All these papers compared several methods and aimed to identify the most accurate overall. Even when using three sets of data at different locations, scales and resolutions, there was no evidence that any particular

technique was consistently better than any other (Thuiller *et al.*, 2003a). Here, an alternative approach that uses different modelling techniques within a single framework has been proposed, where running all of them provides accurate current predictions and reliable future projections. Other modelling techniques could be added to this framework. For instance, genetic algorithms as used by Peterson and co-workers (Stockwell & Peters, 1999; Peterson *et al.*, 2001, 2002a) could be integrated into this framework to cover an even larger spectrum of modelling techniques.

If the aim of a study is to perform only predictions on current data, BIOMOD is useful because it selects the most accurate modelling technique for each species. In this sense, predictions are optimized as compared with other approaches using only one *a priori* modelling technique. Of course, for some species and locations, the different modelling techniques could give exactly the same results, and thus extra calculation time to run

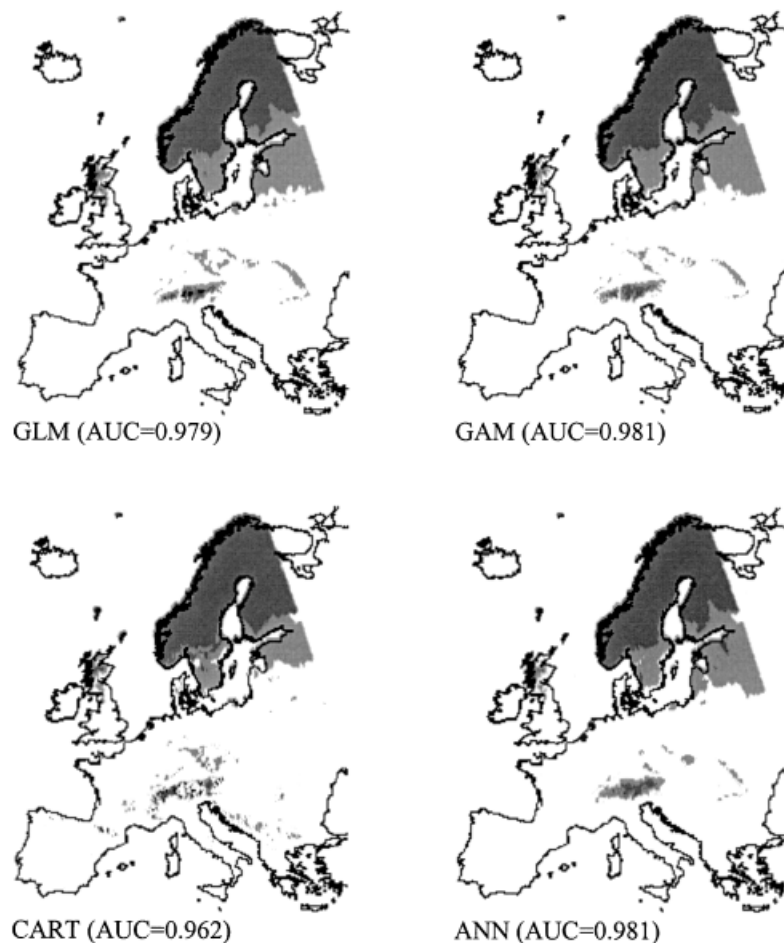


Fig. 5 Comparison of current projected distribution of *Betula nana* and its future potential distribution in 2050. Dark grey colour corresponds to stable suitable sites, grey to loss of suitable sites, black to gain of suitable sites and clear grey to stable unsuitable sites.

different models would be unnecessary (for instance, GLM, GAM and ANN for *B. nana*).

Another advantage of BIOMOD is with respect to projections of species distributions under climate change scenarios. Indeed, it is impossible to gate whether projected future potential distributions are accurate or not, or if one method is better than another, even if expert knowledge could help to assess the reliability of projections. Two alternatives are then possible:

- Either, decide to trust the evaluation process, select for each species the most powerful modelling technique to predict current distributions and project future potential distributions.
- Or, carry out predictions and derive projections of future potential distributions using all the modelling techniques employed (i.e. Figs 2–5) and draw conclusions by comparisons across outputs. These conclusions and validations could be based on expert

knowledge or published studies. A complementary way is to analyse projected future response curves to environmental gradients to identify unrealistic patterns generated by particular modelling techniques (Thuiller *et al.*, in review-a).

Alternatively, the best approach to assess a model's accuracy in the future would be to evaluate models retrospectively to produce projections of the present-day distribution for data gathered in the past. Unfortunately, accurate past data of species and climate are not readily available for Europe at this point.

Projections of potential future distributions also need to be interpreted with caution. If any model is used to project outside the range over which it has been fitted, projections could be less reliable (Thuiller *et al.*, in review-a). Even if models presented in this study are quite accurate and commonly used to assess the impact of global change (Sykes, 2001; Bakkenes *et al.*, 2002; Pearson *et al.*, 2002), particular care should be taken for

species at the edge of their distributions in the study area and for which models can therefore not capture the entire suitable environmental range.

I conclude by proposing to use a large spectrum of methodological approaches (parametric, non-parametric, rule-based or machine learning) to have a wide array of possible future species distributions and assess their reliability using (i) the evaluation methods and (ii) expert knowledge and ecological validation. BIOMOD could also be used to assess the uncertainty of projections of future species distributions. At least three sources of uncertainties are inherent to the modelling process presented in this paper and other published studies (Bakkenes *et al.*, 2002; Berry *et al.*, 2002; Midgley *et al.*, 2002; Peterson *et al.*, 2002b): the uncertainty based on the climate change scenarios, the uncertainty based on the modelling techniques and the uncertainty based on the selection of the threshold to transform probability values into presence-absence form. The framework presented in this paper could measure the prevalence of these three sources of uncertainties and in this sense evaluate the reliability of projections of future species distributions.

Acknowledgments

This research was funded by the European Commission's ATEAM (Advanced Terrestrial Ecosystem Analysis and Modelling) project (EVK2-CT-2000-00075). Species data were kindly supplied by R. Leemans and M.B. Araújo. The author particularly thanks T.D. Mitchell of the Tyndall Centre for Climate Change Research for providing climate and scenario data and S. Zaehle and M. Erhard of Potsdam-Institute for Climate Impact Research for aggregating climate data. The author also thanks Miguel B. Araújo, Lluís Brotons, Sandra Lavorel, Richard Pearson and two anonymous referees for their valuable comments on the manuscript.

References

- Akaike H (1974) A new look at statistical model identification. *IEEE Transactions on Automatic Control*, **AU-19**, 716–722.
- Anon (1999) *S-PLUS 2000 Guide to Statistics*, Vol. 1. MathSoft, Seattle, WA.
- Araújo MB, Williams P (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331–345.
- Austin MP, Meyers JA (1996) Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *Forest Ecology and Management*, **85**, 95–106.
- Bakkenes M, Alkemade RM, Ihle F *et al.* (2002) Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology*, **8**, 390–407.
- Berry PM, Dawson TE, Harrison PA *et al.* (2002) Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. *Global Ecology and Biogeography*, **11**, 453–462.
- Breiman L, Friedman JH, Olshen RA, *et al.* (1984) *Classification and Regression Trees*. Chapman & Hall, New York.
- Brito JC, Crespo EG, Paulo OS (1999) Modelling wildlife distributions: logistic multiple regression vs overlap analysis. *Ecography*, **22**, 251–260.
- Brotons L, Thuiller W, Araújo MB *et al.* (in review) Presence-absence versus presence-only based habitat suitability models for bird atlas data: the role of species ecology and prevalence. *Ecological Modelling*.
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- 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.
- Frescino TS, Edwards Jr. TC, Moisen GG (2001) Modeling spatially explicit forest structural attributes using Generalised Additive Models. *Journal of Vegetation Science*, **12**, 15–26.
- Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modelling of plant species distribution. *Plant Ecology*, **143**, 107–122.
- Hastie TJ, Tibshirani R (1990) *Generalised Additive Models*. Chapman & Hall, London.
- Hill JK, Thomas CD, Huntley B (1999) Climate and habitat availability determine 20th century changes in a butterfly's range margin. *Proceedings of the Royal Society of London B*, **266**, 1197–1206.
- Humphries C, Araujo M, Williams P *et al.* (1999) Plant diversity in Europe: *Atlas Florae Europaeae* and WORLDMAP. *Acta Botanica Fennica*, **162**, 11–21.
- Jalas J, Suominen J (1972–1996) *Atlas Florae Europaeae*. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.
- Lahti T, Lampinen R (1999) From dot maps to bitmaps – *Atlas Florae Europaeae* goes digital. *Acta Botanica Fennica*, **162**, 5–9.
- Leathwick JR (1995) Climatic relationships of some New Zealand forest tree species. *Journal of Vegetation Science*, **6**, 237–248.
- Lehmann A, Overton JM, Leathwick JR (2003) GRASP: generalised regression analysis and spatial prediction. *Ecological Modelling*, **160**, 165–183.
- Manel D, Dias JM, Buckton ST *et al.* (1999) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, **36**, 734–747.
- McCullagh P, Nelder JA (1989) *Generalised Linear Models*. Chapman & Hall, London.
- Midgley GF, Hannah L, Millar D *et al.* (2002) Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. *Global Ecology and Biogeography*, **11**, 445–451.
- Midgley GF, Hannah L, Millar D *et al.* (2003) Developing regional and species-level assessments of climate change impacts on biodiversity: a preliminary study in the Cape Floristic Region. *Biological Conservation*, **112**, 87–97.

- Mitchell TD (2002) A comprehensive set of climate scenarios for Europe *in preparation*.
- Moisen GG, Frescino TS (in press) Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling*, **157**, 209–222.
- Monserved RA, Leemans R (1992) Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, **62**, 275–293.
- Nakicenovic N, Swart R (eds) (2000) *Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge.
- Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.
- Pearce J, Ferrier S (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127–147.
- Pearson RG, Dawson TE (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* (in press).
- Pearson RG, Dawson TP, Berry PM *et al.* (2002) SPECIES: a spatial evaluation of climate impact on the envelope of species. *Ecological Modelling*, **154**, 289–300.
- Peterson AT, Ortega-Huerta MA, Bartley J *et al.* (2002a) Future projections for Mexican faunas under global climate change scenarios. *Nature*, **416**, 626–629.
- Peterson AT, Ortega-Huerta MA, Bartley J, *et al.* (2002b) Biodiversity consequences of climate change in Mexico.
- Peterson AT, Sánchez-Cordero V, Soberón J *et al.* (2001) Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecological Modelling*, **144**, 21–30.
- Root TL, Price JT, Hall KR *et al.* (2003) Fingerprints of global warming on wild animals and plants. *Nature*, **421**, 57–60.
- Rouget M, Richardson DM, Lavorel S *et al.* (2001) Determinants of distribution of six *Pinus* species in Catalonia, Spain. *Journal of Vegetation Science*, **12**, 491–502.
- Stockwell DRB, Peters D (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal Geographical Information Science*, **13**, 143–158.
- Swets KA (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.
- Sykes MT (2001) Modelling the potential distribution and community dynamics of lodgepole pine (*Pinus contorta* Dougl. ex. Loud.) in Scandinavia. *Forest Ecology and Management*, **141**, 69–84.
- Therneau TM, Atkinson EJ (1997) *An introduction to recursive partitioning using the RPART routines*, Technical report, Mayo Foundation.
- Thuiller W, Araújo MB, Lavorel S (2003a) Generalised models versus classification tree analysis: a comparative study for predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* (in press).
- Thuiller W, Brotons L, Lavorel S *et al.* (in review) The effect of area extent in predicting future species distributions. *Ecography*.
- Thuiller W, Howell CA, Midgley GF *et al.* (in review-b) Evaluating species distribution models: methodologies and conservation implications of prevalence and probability thresholds. *Ecological Modelling*.
- Thuiller W, Vaydera J, Pino J *et al.* (2003b) Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology and Biogeography*, **12**, 313–325.
- Vayssières MP, Plant RE, Allen-Diaz BH (2000) Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, **11**, 679–694.
- Venables WN, Ripley BD (2002) *Modern applied statistic with S*. Springer, Berlin.
- Walther G-R, Post E, Convey P *et al.* (2002) Ecological responses to recent climate change. *Nature*, **416**, 389–395.
- Williams P, Humphries C, Araujo MB *et al.* (2000) Endemism and important areas for representing European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belgian Journal of Entomology*, **2**, 21–46.
- Yee TW, Mitchell ND (1991) Generalised additive models in plant ecology. *Journal of Vegetation Science*, **2**, 587–602.

ANNEXE 2.2

Effects of restricting environmental range of data to project current and future species distributions

Thuiller W., Brotons L., Araújo M.B., Lavorel S.

Ecography (2004), sous presse

Effects of restricting environmental range of data to project current and future species distributions

Wilfried Thuiller¹, Lluís Brotons¹, Miguel B. Araújo¹ and Sandra Lavorel^{1,2}

¹Centre d'Ecologie Fonctionnelle et Evolutive. CNRS. 1919 route de Mende, F-34293 Montpellier Cedex 5, France

²Laboratoire d'Ecologie Alpine. CNRS. Université J. Fournier. BP 53X, F-38041 Grenoble Cedex 9, France

Correspondence author: Wilfried Thuiller
E-mail: wilfried.[thuiller@cefe.cnrs-mop.fr](mailto:wilfried.thuiller@cefe.cnrs-mop.fr)

ABSTRACT

We examine the consequences of restricting the range of environmental conditions over which niche-based models are developed to project potential future distributions of three selected European tree species to assess first, the importance of removing absences beyond species known distributions (“naughty noughts”) and second the importance of capturing the full environmental range of species. We found that restricting the environmental range of data strongly influenced the estimation of response curves, especially towards upper and lower ends of environmental ranges. This induces changes in the probability values towards upper and lower environmental boundaries, leading to more

conservative scenarios in terms of changes in distribution projections.

Using restricted data analogous to not capturing the full species’ environmental range, reduces strongly the combinations of environmental conditions under which the models are calibrated, and reduces the applicability of the models for predictive purposes. This may generate unpredictable effects on the tails of the species response curves, yielding spurious projections into the future provided that probability of occurrence is not set to zero outside the environmental limits of the species. Indeed, as the restricted data does not capture the whole of the response curve, projections of future species distributions based of ecological niche modelling may be only valid if niche models are able to approach the complete response curve of environmental predictors.

INTRODUCTION

There is an increasing interest in assessing potential impacts of climate change on species distributions. Amongst modelling techniques available, niche-based models, which are designed to approximate species’ ecological niches (Austin et al. 1990), include some of the most popular methods to project future potential species distributions (Thuiller 2003).

In spite of the popularity of these models (Guisan and Zimmermann 2000, Thuiller et al. 2003a), there are difficulties in projecting species distributions into areas or times different from those used to calibrate models (Lawton 2000, Loehle and LeBlanc 1996, Woodward and Beerling 1997). This is because modelled species’ realised niches capture many factors other than species limits of tolerance to environmental variables (e.g. competition and historical contingency, Araújo et al. 2001, Leathwick and Austin 2001) and these factors are not easily predictable (Loehle and LeBlanc 1996). However, it is possible that increasing the spatial resolution and extent of the studied area might help reducing the importance of unpredictable inter-specific competition in the projections (Beerling et al. 1995).

One of the critical issues for making reasonable projections of species distributions into different spatial or temporal scenarios is to have appropriate

descriptions of species realised niches. Realised niches are commonly approximated by series of curves describing the likelihood of species’ occurrence following a set of environmental predictors using presence/absence or abundance data.

One of the problems with this procedure is that modellers often lack information on the overall species potential or realised distributions. Hence, response curves are often incomplete descriptions of the responses of species to environmental predictors. Different authors have suggested different approaches to tackle this problem. For example, Austin and Meyers (1996) argued that the inclusion of zeros beyond species known distributions (“the naughty noughts”) might perturb the correct estimation of response curves and suggested restricting model calculations within every species known environmental range. Since statistical models can fit extreme zeros reasonably, others authors argued that such restrictions are unnecessary and may lead to ecologically unrealistic response curves (Oksanen and Minchin 2002).

On the other hand, biased availability of environmental conditions are often encountered when modelling distributions of a large number of species over large areas a common feature when assessing for instance biogeographic effects of climate change on

communities (Bakkenes et al. 2002). This is mostly due to restrictions in data accessibility and sampling strategies, to geographic constraints (species at the edge of ranges) or both, preventing the correct capture of the entire environmental range of species (Thuiller et al. 2003b).

In this study, we examine the consequences of restricting the range of environmental conditions over which niche-based models are developed to project potential future distributions of three selected European tree species to assess first, the importance of removing “naughty noughts” and second the importance of capturing the full environmental range of species.

METHODS

Species and climate data: We modelled climate response curves, realised niches and current and future species distributions for three European tree species: *Pinus mugo*, *Salix appendiculata* and *Quercus crenata*. We selected species with distributions restricted to Europe according to Tutin et al. (1964-1993) and which occur primarily in central Europe. This was required because we wanted to maximise the probability of including species’ complete realised niches and a central European location maximises the probability of species occurring in the centre of climatic gradients. Current distributions were taken from the Atlas Florae Europaeae (AFE) (Jalas and Suominen 1972-1996), which uses 2610 UTM 50x50 km grid cells and was digitised by Lahti and Lampinen (1999). Four environmental variables were selected and converted from 0.5° latitude-longitude maps to UTM 50x50 km grid cells. These included: mean ratio of annual actual evapotranspiration over annual potential evapotranspiration (AET/PET), mean annual growing degree days (5°) (GDD), mean temperature of the coldest month (MTC), and mean annual precipitation (MAP) (Michell, pers. comm).

Models: To describe species’ response curves and approximate current realised niches we used generalised additive models (GAM) (Hastie and Tibshirani 1990, for a comparison with other methods, see Thuiller et al. 2003a). We evaluated the fit of the models with a calibration set of 70% randomly chosen points. Then we compared the fit with projections using the 30% remaining set using a threshold independent method, the area under the relative operating characteristic (ROC) curve (AUC, Hanley and McNeil 1982, Pearce and Ferrier 2000). Three models were run for every species according to the following protocol: 1) using all climatic conditions available for the study area; 2) using data restricted to the climate range of species but by setting a limit to include at least 100 observations of the species absence above and below the last presence (for more detail see Austin and Meyers 1996, Austin et al. 1994); 3) using restricted data without zeros spanning beyond the climatic range of species.

The first case corresponded to what is commonly used to project future species distributions (e.g. Bakkenes et

al. 2002, Peterson et al. 2002a, Thuiller 2003). We considered this data as the reference model for comparison with the remaining models. The second case was performed to test if removing naughty noughts could disturb or improve the projected future distributions. The third case was aimed at analysing the impact of not capturing the full environmental range of species and to project distributions outside the environmental limits used to calibrate models. We assume that cutting the data just after the last presence on each gradient would be similar to having species with just the edge of their distributions in the studied area, or to have insufficient sampling points beyond a given location. To compare the predictive accuracy of models, using a same range of environmental data, we evaluated the three models on the evaluation data from the restricted data.

Projected response curves: Response curves were then used to project species distributions on a larger European window using current and future climatic data at finer resolution (10’ grid) (Thuiller et al. in press). Projections for the reference model were carried out simply by using the model calibrated on “all data” and projecting on the new data set. For “restricted + 100”, we followed Austin and Meyers (1996)’ approach, projecting models onto a new data set, but setting probabilities equal to zero above and below the climatic limits used to calibrate models.

Assuming that “restricted” data for a selected species could correspond to a sub-set of the full suitable environmental range of the species, we extrapolated the response curve outside the environmental conditions used to calibrate the model without setting probabilities equal to zero outside the environmental range used to calibrate model. This method is commonly used when researchers predict and project large dataset of species without knowing if they captured the full environmental range of each species (e.g. Bakkenes et al. 2002, Peterson et al. 2002b, Thuiller 2003). Alternatively, as previous projections could lead to less accurate distributions, we also projected models by setting probabilities to zero outside the environmental range used to calibrate model.

Mapping of projected and future distributions: We then mapped current projected species distributions from each model. We also estimated potential future distributions using the third Hadley Global Circulation Model (HadCM3) under one emission scenario (A2) at the same resolution (For details of emission scenarios, see the IPCC’s Special Report on Emissions Scenarios, Nakicenovic and Swart 2000).

RESULTS

The predictive accuracy of models was similar for all models (Table 1).

Calibrated response curves

The shapes of response curves were also similar across the three models, although the removal of zeros beyond the species current climatic (‘restricted’

model) range affected the tail of response curves. We illustrated this by analysing the response curve for the most significant variable among all species, i.e. mean temperature of the coldest month (MTC) (Fig. 1).

For the whole selected species, curves were similar between models but with marked differences in the tails. Left tails for the “restricted” models showed noticeable difference (higher probability values) with the two other models for all species. Response curve for *Salix appendiculata* using “restricted + 100” data had, however, higher probability values than response curve derived from “all data” (Fig. 1).

Projected response curves

Projected response curves showed that the differences, between the approaches of setting probabilities equal to zero outside the environmental range used to calibrate the models and of letting the models estimate the response without any constraint, were important (Fig. 2). The projected response of “restricted data” model without setting probability values equal to zero outside the environmental range used to calibrate models exhibited an strong increase of probability of occurrence for cold temperatures while “all data” model exhibited the inverse trend. This had strong implications for the mapping.

Mapping of projected and future distributions

For instance, projections of *Quercus crenata* using “restricted” models differed noticeably according to the approach used. Setting probability values equal to zero outside the environmental limits used to calibrate models provided similar projections than “restricted + 100” (Fig. 3 D). If probability values outside the calibrated environmental limits were not setting to zero, projections completely over-estimated the observed distribution of *Q. crenata*. This later method showed a strong monotonic increase of probability with a decrease of temperature resulting from a problem of the left tail of the fitted response curve (Fig. 2B) that increased for low values of temperature. So when the response curve was projected to the larger European window, the response curve for low temperature continued to increase while the response curve for “restricted + 100” and “all data” had been set to 0 (Fig. 2B).

Projections of *Quercus crenata* distribution were very similar between “all data” and “restricted + 100” models (Fig. 3 B-C). Projections using “restricted + 100” data and setting probability values equal to zero outside the environmental range used to calibrate models showed more conservative distribution than “all data” model.

Difference between methods to project “restricted” models was exacerbated in the future with spurious projections for models without setting zeros outside the environmental limits of calibrated model (Fig. 4 C-D). Again, “restricted +100” projections were still more conservative than projections using “all data” models (Fig. 4 A-B).

DISCUSSION

We found that restricting the environmental range of data strongly influenced the estimation of response curves, especially towards upper and lower ends of environmental ranges. This has a consequence of first modifying the probability values towards upper and lower environmental boundaries, and second leading to more conservative or liberal projections, depending on the approach.

Effect of restricting the range of environmental conditions on model accuracy

We applied the approach proposed by Austin and Meyers (1996) concerning the effects of naughty noughts to approximate the species’ realised niche in a different context, that is of projecting distribution outside the geographic area used to calibrated models and modelling the future distribution of species under global change. We showed that in terms of accuracy, both “all data” and restricted + 100” were quite similar which demonstrated that removing the naughty noughts does not improve or change the predictive performance of models. The projections of realised niche on 10’ grid and into future climatic conditions demonstrated that removing the naughty noughts to calibrate models and project distribution setting probability values equals to zero outside the environmental limits used, provided more conservative distributions than using all data without restrictions. These results were in total agreement with Austin and Meyers (1996). However, Austin and Meyers (1996) regarded over-predictions as a serious error whereas in a context of global change we tend to regard under-predictions as more serious. When considering using niche-based models to project future distributions, reasonable over-predictions are regarded as existing suitable habitats that species have yet not occupied because of history, constraints to dispersal or other ecological non-equilibrium reasons. From a conservation perspective, such suitable habitats are of interest as they locate the availability of potentially suitable areas, which may then used to plan reintroductions of species or relocate reserves.

Moreover, from a more ecological standpoint, one can see species environmental responses as a gradual continuum along ecological gradients and, as such, it seems more intuitive that species probability of occurrence decreases gradually beyond where it is found rather than assuming a truncated response in which the probability of species occurrence becomes zero beyond the current extent of occurrence.

Effect of restricting the range of environmental conditions beyond the limits used to calibrate models

Restricting data entails several implications and consequences: First, absences are often true absences providing potentially relevant information on species ecology. Using restricted data (similar to not capture the full species’ environmental range) reduces strongly the combinations of environmental conditions under

which the models are calibrated, and reduces the applicability of the model for predictive purposes (Pearson and Dawson 2003). This problem has important implications when future projections of species distributions are sought. In particular, species niche from restricted data sets might be seen as analogous to the modelling of species niche from a limited geographic location not covering the complete range of environmental conditions in which species may occur. We have shown that in such cases, there were two possibilities in projecting models to novel environmental conditions: the first was to project distributions using the calibrated models and second was to project using the calibrated model but setting probabilities equal to zero outside the environmental limits used to calibrate them. The first approach provides liberal projections and second can under-predict the true distribution. The question remains as to which is the best approach: to model species niche from a limited geographic extend (edge of distribution)? Or to project distribution outside the environmental limits used to calibrate the models? In the absence of accurate information on the extent of environmental conditions used to calibrate niche-based models, it might be advisable to project to unknown environmental conditions by using a conservative approach as recommended by Austin and Meyers (1996) in another context.

Secondly, from the perspective of predictive modelling, restricting conditions on which models are calibrated may generate unpredictable effects on the tails of the species response curves, yielding spurious projections into the future (without setting probability values equal to zero outside the environmental limits). Indeed, as the restricted data does not capture the whole of the response curve, projections of future species distributions based of ecological niche modelling may be only valid if niche models are able to approach the complete response curve of environmental predictors. As an example of how considering such constraints into bioclimatic modelling, Pearson et al. (2002) carried out species niche models at the European scale to include the full bioclimatic envelope of the species studied and then downscaled in its application to Great Britain, ensuring that when applied to future climate scenarios the model is not used to extrapolate outside its training data range. Our results support this method as being more suitable to extrapolate species distribution from bioclimatic models because it includes a large part of the range of species distribution and of the environmental combinations where the species currently occur or not.

ACKNOWLEDGEMENTS

This research was funded by the European Commission's ATEAM (Advanced Terrestrial Ecosystem Analysis and Modelling) project (EVK2-CT-2000-00075). Species data were kindly supplied by R. Leemans and M.B. Araújo. The authors thank particularly T.D. Mitchell of the Tyndall Centre for

Climate Change Research for providing climate and scenario data and S. Zaehle and M. Erhard of Potsdam-Institute for Climate Impact Research for aggregating climate data. MBA also thanks FCT (SPRH/BPD/5547/2001) and LB European Community (contract HPMF-CT-2002-01987) for funding.

REFERENCES

- Araújo, M. B., et al. 2001. Would environmental diversity be a good surrogate for species diversity? - *Ecography* 24: 103-110.
- Austin, M. P. and Meyers, J. A. 1996. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. - *For. Ecol. Manage.* 85: 95-106.
- Austin, M. P., et al. 1994. Determining species response functions to an environmental gradient by means of a beta-function. - *J. Veg. Sci.* 5: 215-228.
- Austin, M. P., et al. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. - *Ecol. Monogr.* 60: 161-177.
- Bakkenes, M., et al. 2002. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. - *Glob. Change. Biol.* 8: 390-407.
- Beerling, D. J., et al. 1995. Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surface. - *J. Veg. Sci.* 6: 269-282.
- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. - *Ecol. Model.* 135: 147-186.
- Hanley, J. A. and McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. - *Radiology* 143: 29-36.
- Hastie, T. J. and Tibshirani, R. 1990 *Generalized additive models* - Chapman and Hall, London.
- Jalas, J. and Suominen, J. 1972-1996 *Atlas Florae Europaeae* - The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.
- Lahti, T. and Lampinen, R. 1999. From dot maps to bitmaps - *Atlas Florae Europaeae* goes digital. - *Acta Bot. Fennica* 162: 5-9.
- Lawton, J. H. 2000. Concluding remarks: a review of some open questions. In: Hutchings, M. J., E. John & A. J. A. Stewart (ed.), *Ecological Consequences of Heterogeneity*. Cambridge University Press.
- Leathwick, J. R. and Austin, M. P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. - *Ecology* 82: 2560-2573.
- Loehle, C. and LeBlanc, D. 1996. Model-based assessments of climate change effects on forests: a critical review. - *Ecol. Model.* 90: 1-31.
- Nakicenovic, N. and Swart, R., eds. (2000). *Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*, pp 570. - Cambridge University Press, Cambridge.

- Oksanen, J. and Minchin, P. T. 2002. Continuum theory revisited: what shape are species responses along ecological gradients. - *Ecol. Model.* 157: 119-129.
- Pearce, J. and Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. - *Ecol. Model.* 133: 225-245.
- Pearson, R. G. and Dawson, T. E. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? - *Global Ecol. Biogeogr.* 12: 361-372.
- Pearson, R. G., et al. 2002. SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. - *Ecol. Model.* 154: 289-300.
- Peterson, A. T., et al. 2002a. Future projections for Mexican faunas under global climate change scenarios. - *Nature* 416: 626-629.
- Peterson, A. T., et al. 2002b. Biodiversity consequences of climate change in Mexico. -
- Thuiller, W. 2003. BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change. - *Glob. Change. Biol.* 9: 000-000.
- Thuiller, W., et al. 2003a. Generalized Models versus Classification Tree Analysis: a comparative study for predicting spatial distributions of plant species at different scales. - *J. Veg. Sci.* 14: 669-680.
- Thuiller, W., et al. in press. Do we need land-cover data to model species distributions in Europe? - *J. Biogeogr.*
- Thuiller, W., et al. 2003b. Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). - *Global Ecol. Biogeogr.* 12: 313-325.
- Tutin, T. G., et al. 1964-1993 *Flora Europaea* - Cambridge University Press, Cambridge.
- Woodward, F. I. and Beerling, D. J. 1997. The dynamics of vegetation change: health warnings for equilibrium 'dodo' models. - *Global Ecol. Biogeogr. Lett.* 6: 413-418.

Table 1. Area under the curve (AUC) of the ROC approach describing the accuracy of each model projected on restricted evaluation data. *Salix app.* is the abbreviation of *Salix appendiculata*. AUC is ranging from 0.5 to 1. AUC=0.5: null accuracy and AUC=1: excellent accuracy.

	Models	AUC	AUC
		Evaluation	SdDev
<i>Pinus mugo</i>	“all data”	0.87	0.027
	“restricted + 100”	0.87	0.026
	“restricted”	0.86	0.027
<i>Salix app.</i>	“all data”	0.90	0.022
	“restricted + 100”	0.91	0.022
	“restricted”	0.91	0.022
<i>Quercus crenata</i>	“all data”	0.86	0.074
	“restricted + 100”	0.87	0.070
	“restricted”	0.86	0.063

FIGURE CAPTIONS

Fig. 1. Comparisons of the shapes of the response functions of MTC of the three different models: a) *Pinus mugo*; b) *Salix appendiculata*; and c) *Quercus crenata*.

Fig. 2. Projections of response curve of *Quercus crenata* on larger spectrum of climate data at finer resolution. a) The three models setting probability values equal to zero outside the environmental limits used to calibrate models; b) Same as (a) but without setting probability values equal to zero outside the environmental limits used to calibrate models for the restricted model.

Fig. 3. Spatial distribution of *Q. crenata* projected on the whole Europe (lat-long 10°x10°). a) Observed distribution from Atlas Florae Europaea; b) Projected distribution using “all data” model on current climate; c) Projected distribution using “restricted +100” model on current climate with setting probability values equal to zero outside the environmental limits used to fit models; d) Projected distribution using “restricted” model on current climate with setting probability values equal to zero outside the environmental limits used to fit models; e) Projected distribution using “restricted” model on current climate without setting probability values equal to zero outside the environmental limits used to fit models.

Fig. 4. Future spatial distribution of *Quercus crenata* according to Hadcm3 climate change scenario. Future distribution using: a) “all data” model; b) “restricted + 100” model setting probability values equal to zero outside the environmental limits used to fit model; c) “restricted” model setting probability values equal to zero outside the environmental limits used to fit model; d) “restricted” model without setting probability values equal to zero outside the environmental limits used to fit model.

Fig. 1

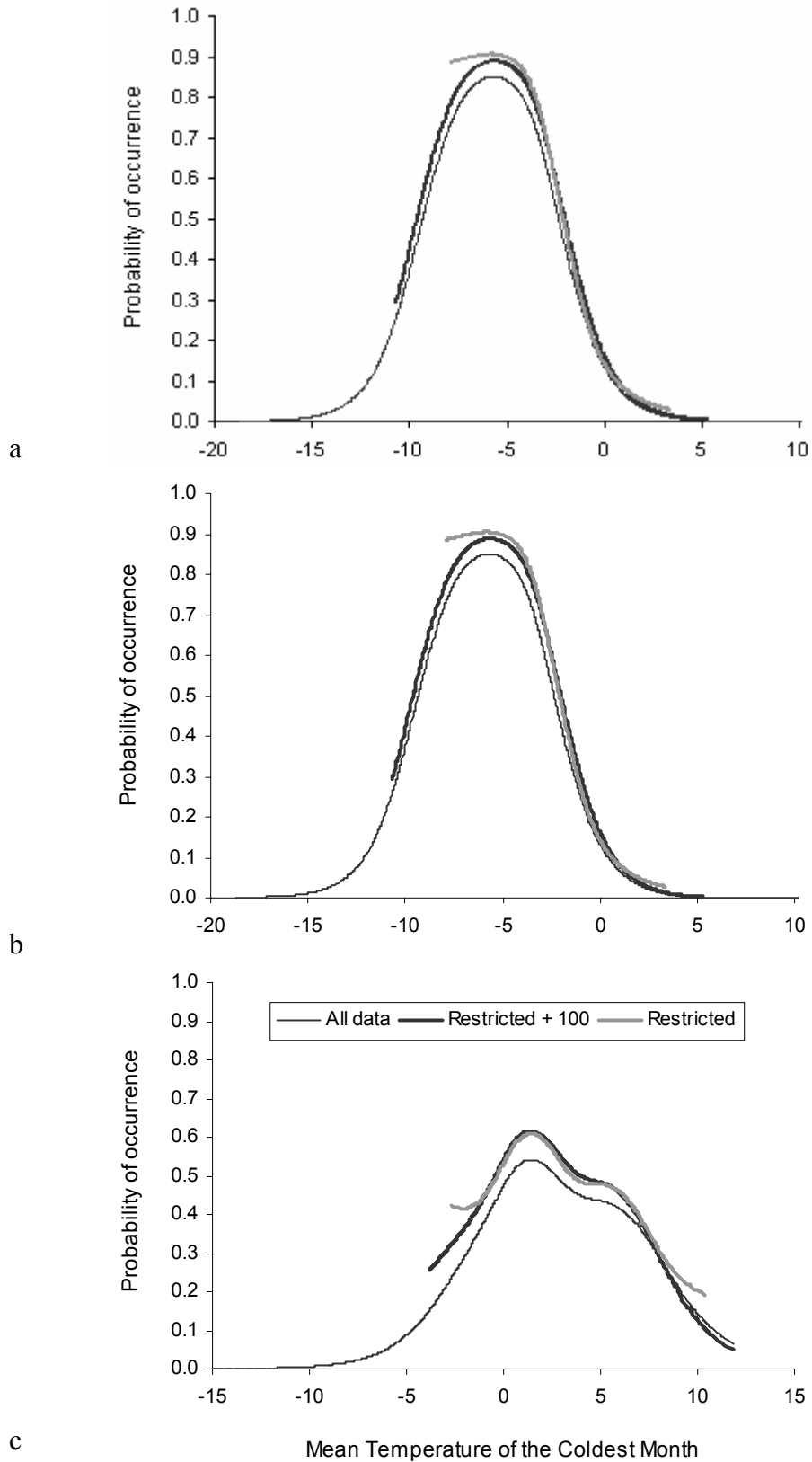
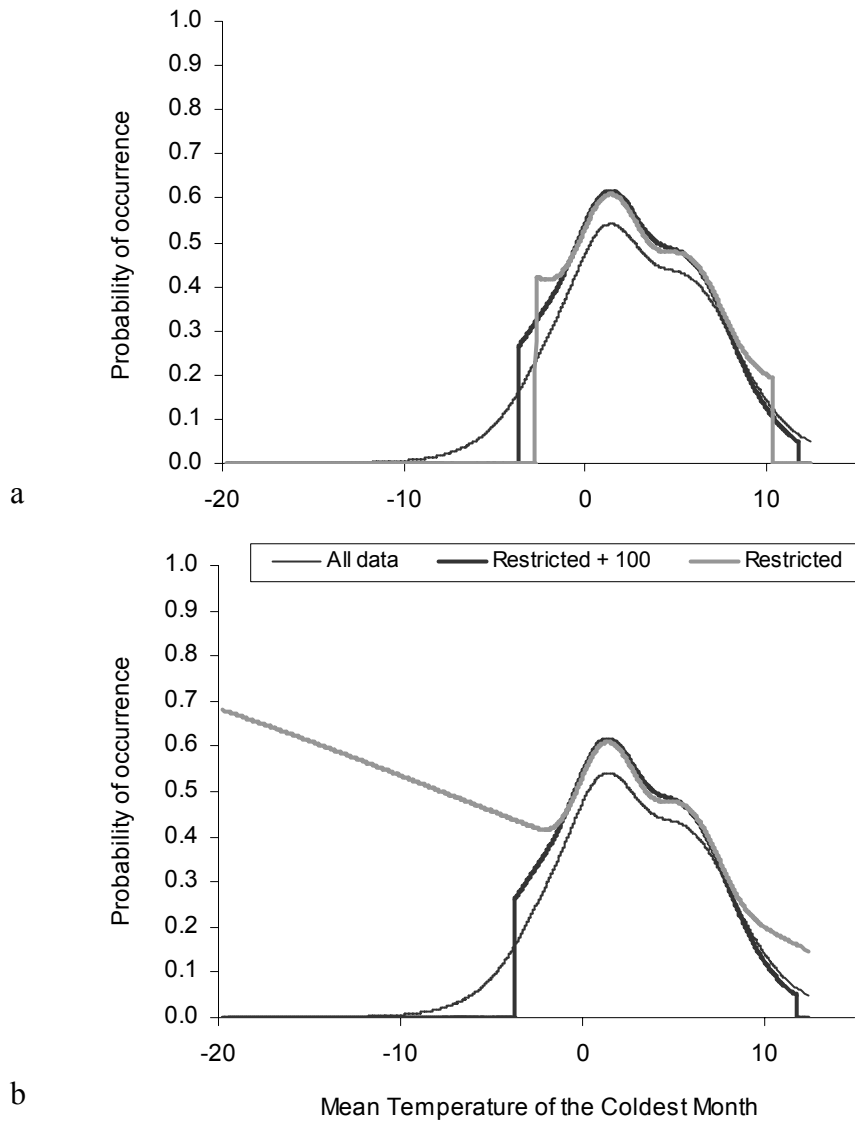


Fig. 2



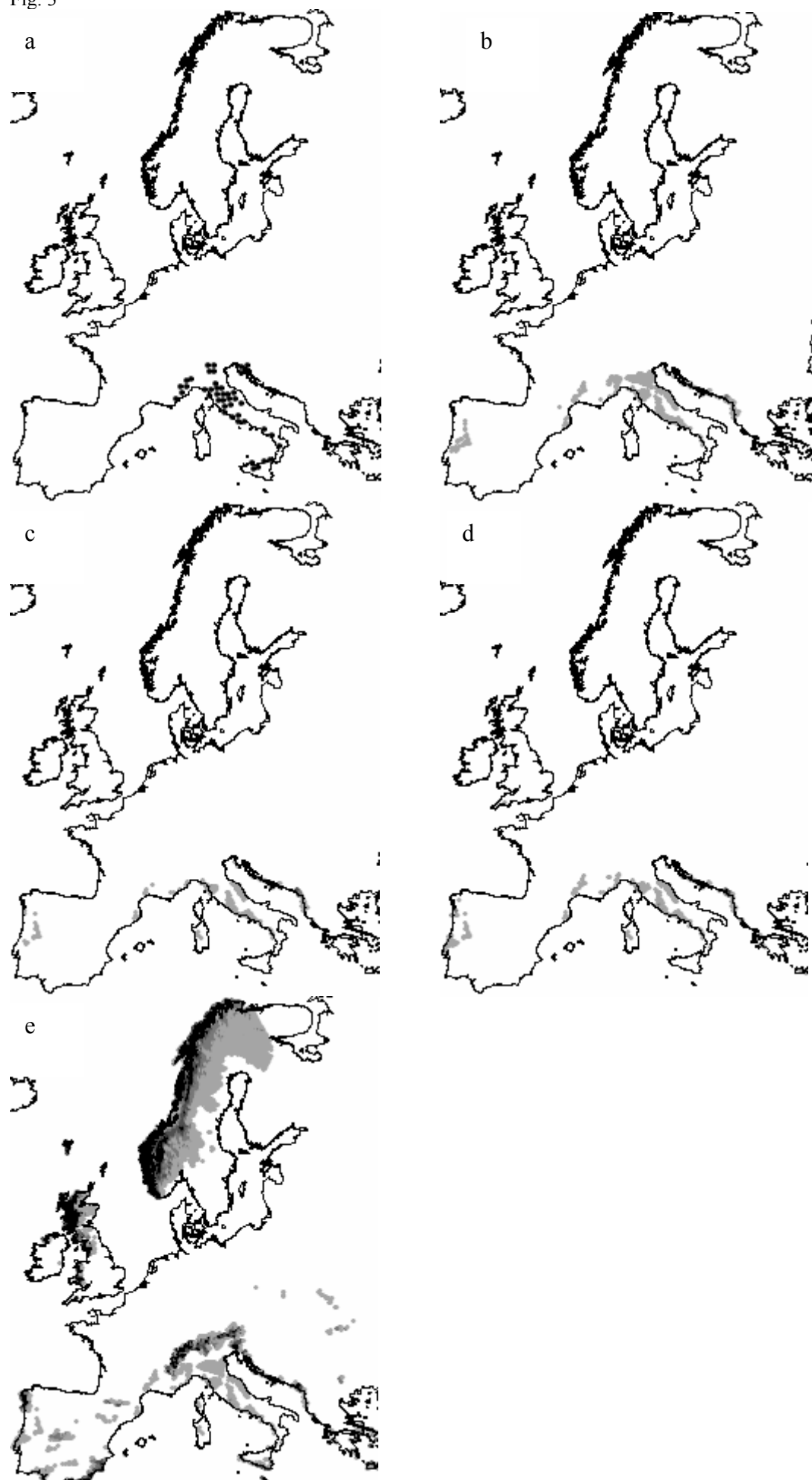
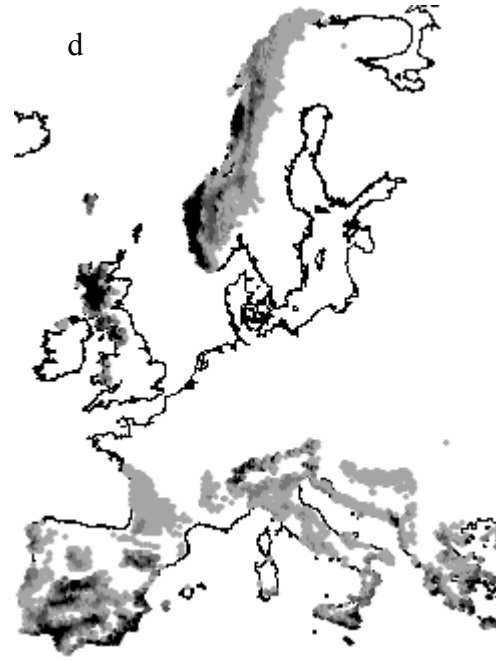


Fig. 4.



ANNEXE 2.3

Developing regional and species-level assessments of climate change impacts on biodiversity: A preliminary study in the Cape Floristic Region

Midgley G.F., Hannah L., Millar D., Thuiller W., Booth A.

Biological Conservation (2003) **112**, 87-97

Developing regional and species-level assessments of climate change impacts on biodiversity in the Cape Floristic Region

G.F. Midgley^{a,b,*}, L. Hannah^b, D. Millar^{a,b}, W. Thuiller^c, A. Booth^d

^a*Climate Change Research Group, Kirstenbosch Research Centre, National Botanical Institute, P/Bag ×7, Claremont 7735, Cape Town, South Africa*

^b*Center for Applied Biodiversity Science, Conservation International, 1919 M St., NW, Washington, DC 20036, USA*

^c*Functional and Evolutionary Ecology Centre, CNRS UPR 9056, Montpellier, France*

^d*Department of Ichthyology, Rhodes University, PO Box 94, Grahamstown 6140, South Africa*

Received 30 May 2002; received in revised form 10 September 2002; accepted 20 September 2002

Abstract

We assess the potential impact of climate change on plant diversity in the Cape Floristic Region (CFR) and its interaction with land transformation that has already occurred in the region. Predictions were made both at the scale of the Fynbos Biome (the dominant vegetation assemblage in the CFR) and for selected Proteaceae species. Bioclimatic modelling identified parts of the biome at particular risk from climate change. Species-level modelling (Generalised Additive Modelling) was done for 28 Proteaceae species selected from areas at high risk of biome loss, revealing individualistic range changes in a pattern broadly consistent with biome modelling results. Most species experienced potential range contractions (17 of 28), of which five showed range elimination. Several species (11 of 28) showed potential range expansions. For species showing range contractions, current land transformation had less impact on future potential ranges than did climate change, because many species ranges shifted to higher altitudes where land transformation is currently less prevalent. Fewer than half of the high-risk species showed overlap between current and future potential range, showing that propagule transport, establishment of species in novel ranges and conservation of landscape linkages will be critical for maintenance of biodiversity. Methods described here provide useful forecasts of potential climate change impacts that could guide conservation responses, but results need cautious interpretation in the light of the many assumptions underlying the techniques used.

© 2003 Elsevier Science Ltd. All rights reserved.

Keywords: Bioclimatic modelling; Climatic niche; Corridor; Geographic range shift; Mediterranean-type ecosystem; Winter rainfall

1. Introduction

Recently documented biotic responses to possible human-induced climate change (Hughes, 2000; Peterson et al., 2002) raise a crucial question for conservationists: Are projected climate changes likely to be a threat to the conservation of biodiversity? Conservation efforts have generally assumed that climate is a constant feature of the environment, and that species distributions are effectively constant in space and time (Cowling, 1999). Both the former and latter assumptions are unjustifiable, as it is now well understood that the earth's

climate has changed significantly and rapidly on time scales of decades to millennia (Broecker, 1999; Zachos et al., 2001); that species have shifted their ranges as climate has changed (Hewitt, 2000; Huntley and Birks, 1983; Parmesan, 1996; Parmesan et al., 1999); and that these range shifts are often individualistic responses to climate change (Graham and Grimm, 1990; Warren et al., 2001) as opposed to wholesale migrations of ecosystems or biomes. Several authors have begun to explore how conservation plans can begin to assess this threat and incorporate these insights into strategies that will be robust to climate change (e.g. Halpin, 1997; Hannah et al., 2002), and some planning to accommodate climate change impacts has been developed for the CFR (Cowling and Pressey, 2001; Cowling et al., 2003; Rouget et al., 2003) and for the winter rainfall

* Corresponding author. Fax: +27-21-797-6903.

E-mail address: midgley@nbict.nbi.ac.za (G.F. Midgley).

Succulent Karoo Biome of South Africa (Cowling et al., 1999a). Nonetheless, explicit adaptations to conservation planning in the face of climate change are in their infancy.

The CFR boasts an exceptionally rich flora with high levels of endemism (Cowling et al., 1989), that may result, in part, from a unique climate and climate history (Cowling et al., 1998; Midgley et al., 2001). The mediterranean-type climate of the CFR may now be changing, characterised by increasing temperatures and reduced rainfall, especially in winter months (Tyson et al., 2002; Wand et al., personal communication). Conservation planning for the region has just begun to account for these significant changes and the impact they may have in concert with other threats (Cowling and Pressey, 2001; Cowling et al., 2003; Rouget et al., 2003).

If species range shifts are the likely dominant species response to future climate change, then spatially explicit planning will be fundamental to estimating the rate and direction of species movements required to ensure retention of sufficient range for their persistence. Modelling can help inform efforts to place land under effective conservation management, either in formally protected areas or in unprotected areas suitable as habitat. These tools must be relevant at regional and even sub-regional scales—the scales at which most practical conservation decisions, such as land acquisitions, are made.

In this paper, we develop a biome- and a species-based approach to assessing the regional impacts of climate change on the future distribution of a major floristic group (Proteaceae). We address a number of questions:

1. What biome-level patterns may be expected under future climate change, and what are the comparative potential range shifts in representatives of a dominant taxon, the Proteaceae, in areas of potential biome contraction?
2. How much has land transformation constrained the potential migration of species in response to climate change?
3. What proportion of species are under severe threat of extinction under the projected climate change scenarios?
4. What are the implications of these patterns for conservation planning?

2. Methods

2.1. Study area and selected species

We modelled the Fynbos Biome (sensu Rutherford and Westfall, 1994) and selected endemic members of

the family Proteaceae within the CFR. The Fynbos Biome is the dominant biome within the Cape Floristic Region. We focused attention on the western parts of the CFR, having established large potential impacts of climate change in this region (Midgley et al., 2003). The study area includes much of the western Cape (between 33–35° S and 18–22° E), a region that has a Mediterranean-type climate with more or less intense summer droughts and wet winters resulting from orographic rainfall and the passage of cyclonic cold fronts associated with westerly winds. Moisture derived from the Indian Ocean and the warm Agulhas Current can be transported onto the southerly coastal plains and southern mountains of the CFR at any time of the year, but this is of minor relevance in our area of interest. The soils of the CFR include mainly sandstone-derived, highly leached and dystrophic lithosols associated with the mountains, duplex soils (sand overlaying heavier subsoils) associated with the coastal platform, and relatively uncommon coastal calcareous soils, often associated with range-restricted plant endemics. The Fynbos Biome is a predominantly sclerophyll shrubland, characterised by the pre-eminence of hard-leaved shrubs (mainly of the species-rich family Proteaceae), narrow- and small-leaved subshrubs and leafless multi-stemmed grass-like growth forms (Taylor, 1978). On the whole, vegetation is fire-prone, with fire recurrence interval of between 4 and 40 years (commonly between 10 and 20 years).

We chose to analyse potential range shifts in Proteaceae species found mainly at low altitudes, with current distributions predominantly in areas of the Fynbos Biome predicted to contract significantly in our biome model.

2.2. Climate scenarios

Future climate scenarios were derived from $\sim 3 \times 3^\circ$ coarse scale projections from the general circulation model HadCM2 for the southern African region interpolated finally to 1×1 minute scale (Schulze and Perks, 1999). The interpolation technique used established relationships between current climate and altitude, topography and continentality to derive finer scale data.

2.3. Vegetation shift modelling

Five climatically derived parameters considered critical to plant physiological function and survival were used to construct climate envelopes for the Fynbos Biome as a whole (e.g. Busby, 1991), and to build Generalised Additive Models (GAMs) of climate constraints for individual species (Yee and Mitchell, 1991). Parameters used were mean minimum temperature of the coldest month (T_{\min}), heat units [HU_{18} , annual sum of daily temperatures ($^\circ\text{C}$) exceeding 18°C], annual

potential evaporation (PE), winter soil moisture days (SMD_{win}), and summer soil moisture days (SMD_{sum} —the latter parameters refer to the number of days for which soil moisture and air temperature were favourable for plant growth). These variables were among those derived for South Africa under present and projected 2050 climatic conditions by Schulze (1997) and Schulze and Perks (1999), and are consistent with variables chosen for similar studies in other regions (Bakkenes et al., 2002).

The minimum temperature of the coldest month is likely to discriminate between species based on their ability to assimilate soil water and nutrients, and continue cell division, differentiation and tissue growth at low temperatures (lower limit), and chilling requirement for processes such as bud break and seed germination (upper limit). Heat units discriminate between species based both on their requirement for a minimum temperature to complete growing cycles (lower limit), and ability to tolerate excess tissue temperature (upper limit). Potential evaporation discriminates through processes related to transpiration-driven water flow through the plant, and xylem vulnerability to cavitation and water transport efficiency. Soil moisture days in

winter and summer discriminate through determining the intensity and duration of seasonal drought stress and growing opportunities, which are especially crucial for seedling recruitment and the survival of this sensitive plant life stage. However, it is important to note that there is little experimental work on local indigenous species to guide in the choice of any bioclimatically limiting variables, and these were chosen as a hypothetical minimum basic set for defining a bioclimatic envelope in South Africa's indigenous flora, in urgent need of testing.

A potential complication ignored by the techniques used here is that of changing atmospheric CO_2 , which can improve the resource-use efficiencies of plants (Drake et al., 1997). However, what little work has been done on Fynbos species response to elevated CO_2 (Midgley et al., 1995, 1999) has revealed neutral effects of this atmospheric change on plant performance in the Proteaceae—furthermore, it can be expected that elevated CO_2 effects on plant growth will be muted in this nutrient-limited system (Stock and Midgley, 1995).

The spatial distribution of the Fynbos Biome as currently mapped was digitised from Rutherford and Westfall (1994), and modelled as described in Midgley

Table 1
Proteaceae species selected for this study, and their key physiognomic and habitat characteristics (after Rebelo, 2001)

Species	Adult height (m)	Habitat	Altitude range (m)
<i>Diastella buekii</i> (Gand.) Rourke	0.15	Moist sands	200–300
<i>Diastella parilis</i> Salisb. ex Knight	0.7	Sandstone foothills	250–500
<i>Diastella proteoides</i> (L.) Druce	0.5	Tertiary and recent sands	0–150
<i>Leucadendron chameleae</i> (Lam.) I. Williams	2.3	Level sandstone sands	150–1000
<i>Leucadendron cinereum</i> (Sol. Ex Aiton) R.Br.	1	Level, sandy soil	15–100
<i>Leucadendron corymbosum</i> P.J. Bergius	2	Level, wet, clay subsoils	100–300
<i>Leucadendron foedum</i> I. Williams	2.5	Sandy soils	30–100
<i>Leucadendron galpinii</i> E. Phillips & Hutch	2–3	Sandy soils	0–200
<i>Leucadendron lanigerum</i> H. Buek ex Meisn. var. <i>lanigerum</i>	1.5	Level clays	180–200
<i>Leucadendron levisanus</i> (L.) P.J. Bergius	2	Damp, sandy soil	0–100
<i>Leucadendron stellare</i> (Sims) Sweet	2	Level, dry sands over clay	30–170
<i>Leucadendron thymifolium</i> (Salisb. ex Knight) I. Williams	2	Sands or gravels over clay	100
<i>Leucospermum arenarium</i> Rycroft	0.75	Deep, white, tertiary sands	120–170
<i>Leucospermum hypophyllocarpodendron</i> subsp. <i>canaliculatum</i> (H. Buek ex Meisn.) Rourke	0.2	Deep, white, recent sands	0–200
<i>Leucospermum muirii</i> E. Phillips	1.5	White tertiary sands	90–260
<i>Leucospermum parile</i> (Salisb. ex Knight) Sweet	1.5	Tertiary and recent sands	30–170
<i>Leucospermum rodolentum</i> (Salisb. ex Knight) Rourke	3	Tertiary and recent sand flats	0–300
<i>Leucospermum tomentosum</i> (Thunb.) R.Br.	0.4–1	Tertiary and recent sand flats near sea	0–80
<i>Protea scolymocephala</i> (L.) Reichard	0.5–1.5	Sandy flats and coastal lowlands	0–400
<i>Serruria adscendens</i> (Lam.) R.Br.	0.5–1	Sandstone soils	0–950
<i>Serruria brownii</i> Meisn.	0.3–0.5	Granite soils, shales and heavy sands	50–250
<i>Serruria candicans</i> R.Br.	0.4–0.8	Granite and sandy soils	60–160
<i>Serruria cyanoides</i> (L.) R.Br.	0.3–0.5	Sands	0–150
<i>Serruria decipiens</i> R.Br.	0.6–1	Sandstone and sandy soils	0–250
<i>Serruria decumbens</i> (Thunb.) R.Br.	0.1	Sandstone soils	150–310
<i>Serruria fucifolia</i> Salisb. ex Knight	0.8–1.5	Sandstone and sandy soils	150–920
<i>Serruria linearis</i> Salisb. ex Knight	0.4–0.8	Sands	120–190
<i>Serruria trilopha</i> Salisb. ex Knight	0.3–0.8	Sands	50–310

et al. (2003). Areas in which there were large differences between present and future modelled biome extent were deemed to be at high risk.

Based on biome-level projections of areas of high climate change impact, we selected for study a set of 28 Proteaceae species currently distributed in lowland regions (see Table 1). We used the Protea Atlas Database (Rebello, 2001) which contains records of both species presence and absence at more than twenty thousand points within the CFR. These data allow the application of GAMs to derive bioclimatic relationships for the selected species (Hastie and Tibshirani, 1990). In addition to the climatic variables described earlier, we included in the species models three soils variables—fertility, sand and clay content, as defined and mapped by Schulze (1997). This is a minimum set of soil factors that is likely to influence plant performance through nutrient availability, and the impacts of soil texture on soil water availability.

The GAM approach maps out areas of differing probability of encountering the species modelled, ranging from 0 to 1. For the purposes of this paper, we chose probability levels of 0.5 and higher to define the bioclimatic and edaphic constraints for each species. This conservative cut-off reduces the risk of identifying areas as suitable for the species even if they are not—it therefore reduces the identification of false positives, or making errors of commission.

While GAMs do not have the same properties as models based on ordinary least squares regression, it is possible to derive a good indication of the performance of the model from a statistic which is analogous to R^2 , but calculated as $1 - (\text{residual deviance}/\text{null deviance})$, referred to here as “explained deviance”. If the model did not improve on a null (random) model for a given species, that species was assumed to have no current bioclimatic or edaphic limitation and its response to climate change not considered further (two species).

Generalised Additive Models for the selected species had adequate to good explanatory power (explained deviance ranged from 0.5 to 0.8, Table 2) for the 28 species modelled in detail. The relatively high coefficients obtained indicate that the species are likely to be climate-constrained, and that changes in the regional climate will have significant impacts on them. The 28 species selected were chosen by statistical screening of 33 candidate species. Two of the original 33 candidate species were excluded from the analysis due to poor statistical fit to null, and three were omitted from further analysis because the models could not identify any part of the study area in which they had a probability of occurrence greater than 0.5 in either the present or the future.

The impact of land transformation on future range was assessed by screening out areas currently mapped as transformed (CSIR, 1999), and assuming that these are unsuitable as habitat for the species.

Table 2
Explained deviance for the Generalised Additive Models

Species	Explained deviance
<i>Diastella buekii</i>	0.77
<i>Diastella parilis</i>	0.74
<i>Diastella proteoides</i>	0.68
<i>Leucadendron chamelaea</i>	0.60
<i>Leucadendron cinereum</i>	0.65
<i>Leucadendron corymbosum</i>	0.62
<i>Leucadendron foedum</i>	0.57
<i>Leucadendron galpinii</i>	0.69
<i>Leucadendron lanigerum</i> var. <i>lanigerum</i>	0.58
<i>Leucadendron levisanus</i>	0.66
<i>Leucadendron stellare</i>	0.56
<i>Leucadendron thymifolium</i>	0.70
<i>Leucospermum arenarium</i>	0.71
<i>Leucospermum hypophyllocarpodendron</i> subsp. <i>canaliculatum</i>	0.65
<i>Leucospermum muirii</i>	0.70
<i>Leucospermum parile</i>	0.77
<i>Leucospermum rodolentum</i>	0.63
<i>Leucospermum tomentosum</i>	0.79
<i>Protea scolymocephala</i>	0.51
<i>Serruria ascendens</i> var. <i>decepiens</i>	0.54
<i>Serruria brownii</i>	0.58
<i>Serruria candicans</i>	0.72
<i>Serruria cyanoides</i>	0.58
<i>Serruria decepiens</i>	0.60
<i>Serruria decumbens</i>	0.68
<i>Serruria fucifolia</i>	0.68
<i>Serruria linearis</i>	0.68
<i>Serruria trilopha</i>	0.61

No attempt was made to refine species-specific models by variable selection, thus the models are termed “saturated” models. The GAM procedure used recorded presence/absence data from the Protea Atlas Database (Rebello, 2001) for each of the species listed below as dependent variables, and physiologically significant substrate and bioclimatic variables (Schulze, 1997) as independent variables.

3. Results and discussion

3.1. Regional-level assessment

Biome-level modelling shows that the Fynbos Biome stands to lose significant areas near its northerly (equatorward) limits, especially in the coastal forelands and inland plains along the west coast (see Fig. 1) in response to projected climate change. The biome envelope suggests future contraction southwards onto the mountains of the Cape Fold Belt. Plains and slopes at lower altitudes along the west coast and northern borders of this mountain belt do not retain suitable bioclimates for vegetation of the Fynbos Biome. These areas are therefore of special conservation concern, as it is unlikely that their vegetation and habitats have experienced this extent and rapidity of warming and desiccation in the recent geological past.

The projected future temperature change represents a significant deviation from conditions in the CFR since the last glacial maximum (LGM), and probably since the inception of the Pleistocene. Modelled mean annual regional temperatures were on the order of 3.7 °C cooler during the LGM than today (Midgley and Roberts 2001; Midgley et al., 2001), and future projections suggest roughly 1.8 °C warmer mean annual regional temperatures, which is clearly a significant change within a very short time frame of 50 years. Future projected changes in precipitation regime are also without recent precedent.

Future predictions for the western Cape are for somewhat drier conditions, which will be exacerbated by higher air temperatures, together resulting in more intense plant water stress. This also is a change unprecedented in the past 20,000 years or more. In all likelihood, moisture availability during glacial conditions was significantly enhanced, due to more frequent and predictable arrival of frontal rainfall, which probably impinged on the west coast further north than is the case today, and resulted in more northerly distribution of currently mesic-associated groups such as the family Restionaceae (Shi et al., 1998), and proven greater prevalence of mesic forest elements north of the current northern limits of the CFR (Cowling et al., 1999b; Parkinson et al., 2000). From the earlier, it is clear that the CFR faces significantly warmer and more arid conditions that may well be without precedent in the recent evolutionary history of many elements of the extant flora. As such, climate change could represent a significant threat to the persistence of these biota.

3.2. Species-level assessment

The 28 species modelled (Table 1) were selected from the area of projected biome loss, mainly low-altitude and coastal regions. They showed a diverse array of range change responses to the climate change scenario applied (see Table 3), and these provide some indication of risk due to climate change. Three major risks face species in a changing climate—range elimination, range reduction and range shift.

Range elimination: five of the 28 species (18%) are projected to have no suitable geographic range under the future scenario. These can be considered species most at risk of extinction under the given climate scenario, but cautious interpretation of this result is necessary. Modelled losses of range, even if they are derived from accurate knowledge of bioclimatic limits of species, may not represent immediate extinction. It is possible, for example, that established adults could persist in regions where climate has changed beyond their modelled limits because limits may be determined by the sensitivity of establishing seedlings. This effect would lead to a differential elimination of species dependent on seeds for post-fire regeneration (re-seeders), but allow plants which survive fire (resprouters) to persist. The resprouting habit is not evenly spread among the dominant genera in Fynbos (Le Maitre and Midgley, 1992)—this effect could therefore result in significant shifts in species composition and ecosystem function. The bioclimatic modelling approach fails to recognise this and other aspects of life history and population dynamics, which limits its ability to predict the detailed evolution of climate change effects in nature.

Range reductions: the projected climate change scenario reduces the modelled range sizes of 12 of the

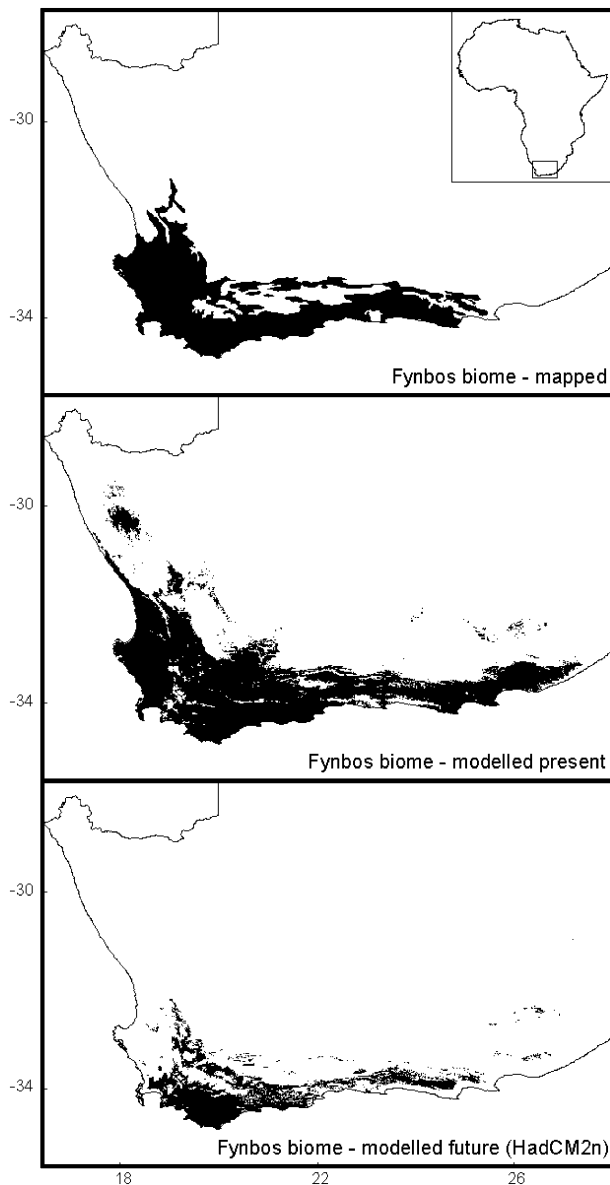


Fig. 1. Current mapped Fynbos Biome (upper panel, after Rutherford and Westfall, 1994), and the modelled extent of the biome under current (middle panel) and future (~2050) climate conditions (lower panel), the latter based on climate change projections for the region generated by the GCM HadCM2n.

28 species modelled (42% of the species, Table 3). Among these species the mean loss of range is 84% (median loss 90%), with the range of values falling between 27 and 99% of current range.

Although range losses predominate, a significant proportion (39%) of species show potential gains in range. This echoes the results of Warren et al. (2001) and Peterson et al. (2002) who have found that species display a range of responses to climate change, both positive and negative. Pimm (2001) has pointed out the difficulty this diversity of response poses for predictive modelling efforts.

Clearly, some species stand to gain from the climate change, but realisation of this theoretical new range depends on species dispersal capability, population size, fecundity and ability to penetrate new habitats. Experimentally derived information on seed dispersal distances, fecundity and establishment abilities are sorely lacking in indigenous Cape vegetation, and compromises the predictive power of this kind of modelling

approach. Furthermore, the coarse soil fertility and texture classification we have used for this study may over-estimate the potential ability of species to occupy new ranges, and our range shift results could be seen as optimistic values.

Land transformation has a clear negative impact on potential range size (Table 3). Under current climate conditions, an average of 55% of potential range for the test species has been transformed, but perhaps surprisingly, this figure drops to only 27% under future climate conditions. This is the result of species ranges shifting to higher altitudes where land transformation is minimal (Fig. 2). These results do not account for future increases in land transformation, and are also subject to the modelling technique's assumptions discussed earlier.

Range shifts: twenty-three of the 28 species modelled showed range shifts (the remaining five species suffered range elimination). Thirteen of these 23 species had no geographic overlap between current and future projected ranges.

Table 3

Current modelled and future projected range sizes (total number of minute×minute squares with modeled probability of occurrence > 0.5) with and without taking into account the impact of land transformation

Species	Modelled range size			
	Current		2050	
	Without transformation	With transformation	Without transformation	With transformation
<i>Diastella buekii</i>	12	4	3	3
<i>Diastella parilis</i>	15	9	0	0
<i>Diastella proteoides</i>	63	26	<i>114</i>	<i>53</i>
<i>Leucadendron chamelaea</i>	45	16	1	1
<i>Leucadendron cinereum</i>	154	55	<i>1317</i>	<i>538</i>
<i>Leucadendron corymbosum</i>	66	35	5	3
<i>Leucadendron foedum</i>	1	1	83	73
<i>Leucadendron galpinii</i>	218	112	1	1
<i>Leucadendron lanigerum</i> var. <i>lanigerum</i>	352	71	43	24
<i>Leucadendron levisanus</i>	17	3	<i>308</i>	<i>219</i>
<i>Leucadendron stellare</i>	103	11	<i>366</i>	<i>147</i>
<i>Leucadendron thymifolium</i>	15	0	<i>257</i>	<i>206</i>
<i>Leucospermum arenarium</i>	27	10	5	0
<i>Leucospermum hypophyllocarpodendron</i> subsp. <i>canaliculatum</i>	263	114	30	19
<i>Leucospermum muirii</i>	45	26	0	0
<i>Leucospermum parile</i>	81	39	1	1
<i>Leucospermum rodolentum</i>	1276	707	3	3
<i>Leucospermum tomentosum</i>	513	319	<i>372</i>	<i>216</i>
<i>Protea scolymocephala</i>	2	0	<i>35</i>	<i>24</i>
<i>Serruria adscendens</i> var. <i>decipiens</i>	9	8	<i>165</i>	<i>131</i>
<i>Serruria brownii</i>	1	0	0	0
<i>Serruria candicans</i>	22	22	9	9
<i>Serruria cyanoides</i>	0	0	2	2
<i>Serruria decipiens</i>	149	78	0	0
<i>Serruria decumbens</i>	0	0	<i>18</i>	<i>17</i>
<i>Serruria fucifolia</i>	744	377	21	19
<i>Serruria linearis</i>	28	15	0	0
<i>Serruria triloba</i>	0	0	<i>108</i>	<i>102</i>

Italicised values for 2050 highlight range size gains.

The vector of potential species range shifts (derived by connecting centroids of the current and future species ranges) suggest a general south-eastward displacement in response to projected climate change, although some species show large eastward shifts (Fig. 3). The vector distances linking current and future population centroids range from 8 to 483 km (Table 4). These distances reflect the distance the main body of a population must traverse to be maintained.

A significant proportion of Fynbos species possess directed dispersal syndromes with short potential distances, such as myrmecochory (Le Maitre and Midgley,

1992). Many of the high-risk lowland species modelled here are ant-dispersed, and cannot be expected to maintain the pace of dispersal indicated by these centroid distances in a timeframe of decades. Interestingly, habitat loss due to existing land transformation does not significantly alter either the distance or direction of range displacement (Fig. 3, Table 4).

Distances between the edge of current distributions and the nearest edge of a future potential range (separation distances) provide a better assessment of the risk of species extinction than do the centroid–centroid distances (range shifts). Separation distances are significantly less than 25 km in most cases and are normally only marginally increased if land transformation is taken into account (Table 4).

Ten species show overlap between their current and future potential ranges, and can be considered least at risk under this climate scenario. However, this number drops to eight when land transformation is considered (Table 4), showing that habitat destruction increases the risk from climate change by accentuating the need for dispersal in a subset of species.

It is noteworthy that even among the species whose ranges are projected to expand, the majority are obligate range-shifters (no overlap between present and future ranges). These species must migrate across considerable distances to avoid extinction, despite the fact that their ranges are projected to increase. This indicates that range expansion is not a reliable index of range overlap or species' vulnerability to climate change. Land transformation is certain to be critical in determining potential landscape linkages for those species without range overlap, but this aspect requires analysis beyond the scope of this paper.

Taken together, and taking cognisance of the assumptions underlying the modelling techniques used here, the majority of the species modelled seem to be at risk of range size reduction and even extinction under the given climate change scenario. Five species have zero future potential range, six lose more than 90% of their range, and a further four are required to migrate more than 10 km between their current and future ranges. These 15 of the 28 species modelled (53%) appear to be at highest risk of extinction under this climate change scenario.

3.3. Implications for conservation planning in the region

These first phase results indicate several important conclusions for the region and several for conservation planning in other areas. At the regional scale, it is clear that west coast lowlands of the Fynbos biome are most vulnerable to climate change impacts. This region is home to a considerable biodiversity (Cowling, 1992), but is relatively poorly conserved (Cowling et al., 1999c; Younge, 2000). Considerable species range shifts forecast

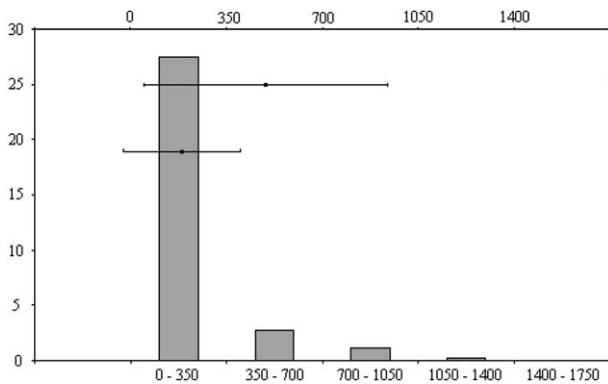


Fig. 2. The current altitudinal distribution of transformed land in the study region (Columns, lower X-axis), and the present and future (~2050) altitudinal distribution of the 28 species modelled in this study (Filled circles and error bars, Upper X-axis).

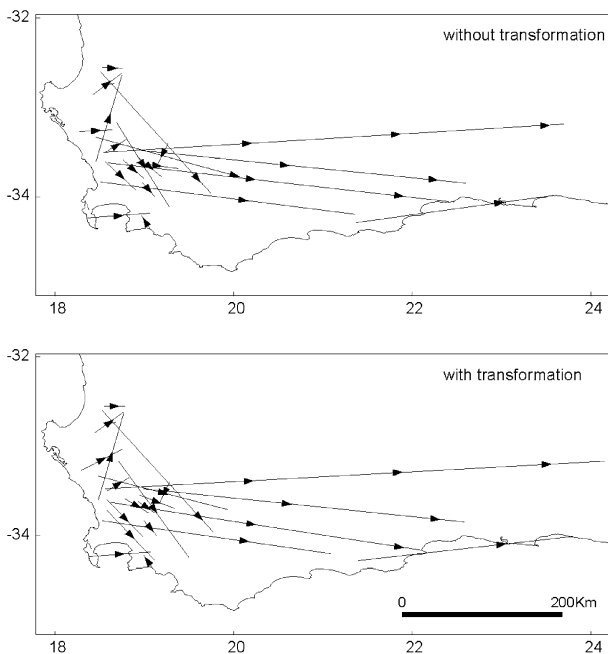


Fig. 3. Modelled range shifts in response to climate change, required for species to retain their optimum bioclimatic range. Arrows join the geographic centroids of the current and future (~2050) ranges of modelled species, without (upper panel) and with (lower panel) the effects of land transformation.

Table 4

Shortest distance (km) between the species' present distributional range and potential future range (nearest edge to edge), with and without taking into account the impact of land transformation (E indicates species has no future range, i.e. it goes extinct under the climate change scenario, and 0 km indicates overlap between current and potential future range)

Species	Centroid-centroid (km)		Closest edge-edge (km)	
	Without transformation	With transformation	Without transformation	With transformation
<i>Diastella buekii</i>	22	22	15	15
<i>Diastella parilis</i>	E	E	E	E
<i>Diastella proteoides</i>	49	59	0	5
<i>Leucadendron chamelaeae</i>	28	28	8	8
<i>Leucadendron cinereum</i>	154	142	0	0
<i>Leucadendron corymbosum</i>	36	40	0	0
<i>Leucadendron foedum</i>	40	40	0	0
<i>Leucadendron galpinii</i>	206	224	7	7
<i>Leucadendron lanigerum</i>	33	31	0	0
var. <i>lanigerum</i>				
<i>Leucadendron levisanus</i>	267	239	0	3
<i>Leucadendron stellare</i>	119	141	8	17
<i>Leucadendron thymifolium</i>	357	328	11	38
<i>Leucospermum arenarium</i>	190	190	157	157
<i>Leucospermum hypophyllocarpodendron</i>	483	523	2	6
subsp. <i>canaliculatum</i>				
<i>Leucospermum muirii</i>	E	E	E	E
<i>Leucospermum parile</i>	29	29	22	22
<i>Leucospermum rodolentum</i>	10	10	5	5
<i>Leucospermum tomentosum</i>	35	50	0	0
<i>Protea scolymocephala</i>	111	111	21	21
<i>Serruria adscendens</i> var. <i>decipiens</i>	19	17	0	0
<i>Serruria brownii</i>	E	E	E	E
<i>Serruria candicans</i>	340	340	0	0
<i>Serruria cyanoides</i>	7	7	0	0
<i>Serruria decipiens</i>	E	E	E	E
<i>Serruria decumbens</i>	65	63	2	2
<i>Serruria fucifolia</i>	22	22	2	2
<i>Serruria linearis</i>	E	E	E	E
<i>Serruria trilopha</i>	101	80	5	22

in this area suggest that sympathetic management from landowners will be important for maintaining corridors for dispersal of these biota. Detailed modelling of montane species is necessary to confirm that their level of risk is lower than for the mainly lowland species modelled here.

Spatial overlap between current and future ranges is present in a substantial portion of species, even in this high-risk group. Ten species showed overlap between present and future ranges, and eight of these ten maintained overlap even when land transformation was considered. Protected area planning can be directed at these areas with relatively straightforward adaptations of existing systematic protected areas planning tools (reserve selection algorithms).

Two counter-intuitive results emerged. First, range expansion was present in some species with substantial populations in the area projected to experience biome loss. This supports the principle that species, rather than communities, are the unit of response to climate change.

Paleoecological studies have shown community disassociation and species individualistic range shifts in periods of past climate change (Huntley and Birks, 1983). Other assessments of the biotic impact of future climate change have shown similar species individualistic patterns (Bakkenes, 2002; Peterson et al., 2002).

Second, climate change had a greater impact on projected ranges than did current land transformation. The direction of range shifts upslope and the pattern of reduced upland land transformation combine to reduce the effect of habitat loss in future ranges. These results show the importance of species-level modelling and indicate that climate change may be critical to the medium- and long-term success of conservation efforts in the region. As land transformation and climate change are both accelerating, the combined impact of future transformation and climate change may be much larger than indicated here. A conservation window exists, in which near-term actions are much more likely to be feasible and cost-effective.

3.4. Implications for conservation strategies

This study presents unique insight into the relation between modelled biome changes and possible underlying species range shifts. Comparison of biome modelling and species modelling has seldom been attempted. Most studies focus either on global biome modelling (e.g. Prentice et al., 1992; Cramer et al., 2001) or on regional species modelling (e.g. Peterson et al., 2002) and seldom, on biome and species modelling at the same location and scale.

In the CFR, species range shift modelling results support the broad conclusions of biome modelling, but with species-specific characteristics important for conservation. Most species lose range from the area of modelled biome loss. The median range loss within the area of biome loss is 97.8%. Nearly half of the species modelled lose all of their range within the area of biome loss. Yet a substantial number of species expand their ranges within the area of biome loss. Six of 28 species (21%) gain range within the area of projected biome loss. Three of these species more than double their range. An additional three species expand their ranges outside the area of projected biome loss. The proportion of range expanders and species with total loss of range was about the same even in those species that are endemic or near-endemic (>90% of range) to the area of biome loss. Of the 14 endemic and near-endemic species, 21% showed range expansion, while 29% lost all range.

Thus, the area of biome loss indicates an area of wholesale range loss and re-arrangement, but it does not indicate direct climate-driven extinctions for many species. The implication for conservation strategies is that areas of modelled biome loss are priorities for species-level modelling, landscape analysis, and attempts to refine both regional climate predictions and modelling techniques. These are *not* indicated as conservation write-offs, even ignoring the model assumptions discussed above. Even those species (5 of 28) modelled to have no future range at all may persist for sometime in unfavourable climatic conditions or in micro-refuges.

Finer-scale modelling can reveal possible micro-refuge locations. For other species, coordinated creation of new protected areas and landscape linkages across the region will be key to allowing potential future range to be occupied (Hannah et al., 2002). Conservation strategies must expand protected areas efforts and encompass larger landscape conservation efforts. A new, more complex conservation paradigm is required.

Such strategies have been termed Climate Change–Integrated Conservation Strategies (CCS) (Hannah et al., 2002). A CCS begins with climatic and range shift modelling, which in turn informs four later stages. The five elements of a CCS are:

1. Regional Modelling;
2. Protected Areas Revision;
3. Matrix Management (creation of landscape linkages);
4. Regional Coordination of Conservation; and
5. Global Transfer of Resources.

Biome modelling may be a useful first-cut assessment in the regional modelling portion of a CCS. However, the results of this study show that species-level modelling is essential for detailed conservation planning. Range elimination, range contraction and range shifts are all possible climate change outcomes which must be considered both independently and in concert.

Protected areas planning and management must be expanded based on modelling results. These efforts can begin by focusing on areas in which present and future ranges overlap. The CFR modelling shows that even in areas projected for major bioclimatic changes (biome collapse) some species may still show range overlap. A general strategy is to use site selection algorithms to plan protected areas and conservation management in those areas where present and future ranges are projected to overlap. These sites will be adequate to meet minimum range (or population) requirements for some species. For other species, a combination of present range and potential future range will be required to meet conservation targets.

For the remaining species whose ranges do not overlap (the majority in this study), management of the matrix of land uses between protected areas is essential for persistence. Species-level modelling is a major asset in this planning process. Information on the direction and magnitude of range shifts can be used to design landscape connectivity of habitats useful to each target species. This can ultimately lead to a matrix of land uses between protected areas that allows complex multiple range shifts to occur.

Regional coordination of protected areas planning and matrix management efforts will be necessary across political divisions and administrative jurisdictions for such landscape-based approaches to be effective. In the Cape, this coordination is within sub-divisions within a single country. In other regions, coordination across international boundaries will be equally important.

International transfer of resources is necessary to underwrite the modelling and research efforts, additional protected areas, landscape conservation management and conservation coordination required in Climate Change–Integrated Conservation Strategies. Cooperation among international and national conservation groups has initiated this support for the initial stages of modelling in the CFR. Major new funding mechanisms are called for to extend this support to other CCS elements in the Cape and to developing CCS in other regions of the world.

Acknowledgements

We acknowledge with thanks comprehensive funding from the Centre for Applied Biodiversity Science, Conservation International, Washington, DC, and helpful and constructive comments from Richard Cowling and Bob Pressey on an earlier draft of this paper.

References

- Bakkenes, M., Alkemade, J.R., Ihle, F., Leemans, R., Latour, J.B., 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.
- Broecker, W.S., 1999. What if the conveyor were to shut down? Reflections on a possible outcome of the great global experiment. *GSA Today* 9, 1–5.
- Busby, J.R., 1991. A bioclimatic analysis and prediction system. In: Margules, C.R., Austin, M.P. (Eds.), *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. CSIRO, Australia, pp. 64–68.
- Cowling, R.M., 1992. *The Ecology of Fynbos: Nutrients, Fire and Diversity*. Oxford University Press, Cape Town.
- Cowling, R.M., 1999. Planning for persistence—systematic reserve design in southern Africa's Succulent Karoo desert. *Parks* 9, 17–30.
- Cowling, R.M., Pressey, R.L., 2001. Rapid plant diversification: planning for an evolutionary future. *Proceedings of the National Academy of Sciences* 98, 5452–5457.
- Cowling, R.M., Gibbs Russell, G.E., Hoffman, M.T., Hilton-Taylor, C., 1989. Patterns of plant species diversity in southern Africa. In: Huntley, B.J. (Ed.), *Biotic Diversity in Southern Africa. Concepts and Conservation*. Oxford University Press, Cape Town, pp. 19–50.
- Cowling, R.M., Rundel, P.W., Desmet, P.G., Esler, K.J., 1998. Extraordinarily high regional-scale plant diversity in southern African arid lands: subcontinental and global comparisons. *Diversity and Distributions* 4, 27–36.
- Cowling, R.M., Pressey, R.L., Lombard, A.T., Desmet, P.G., Ellis, A.G., 1999a. From representation to persistence: requirement for a sustainable reserve system in the species rich mediterranean-climate deserts of southern Africa. *Diversity and Distributions* 5, 1–21.
- Cowling, R.M., Cartwright, C.R., Parkington, J.E., Allsopp, J.C., 1999b. Fossil wood charcoal assemblages from Elands Bay Cave, South Africa: implications for Late Quaternary vegetation and climates in the winter-rainfall fynbos biome. *Journal of Biogeography* 26, 367–378.
- Cowling, R.M., Pressey, R.L., Lombard, A.T., Heijnis, C.E., Richardson, D.M., Cole, N., 1999. Framework for a Conservation Plan for the Cape Floristic Region (IPC Report 9902) submitted to WWF, SA.
- Cowling, R.M., Pressey, R.L., Rouget, M., Lombard, A.T., 2003. A conservation plan for a global biodiversity hotspot—the Cape Floristic Region, South Africa. *Biological Conservation* 112, 191–216.
- Cramer, W., Bondeau, A., Woodward, F.I., Prentice, I.C., Betts, R.A., Brovkin, V., Cox, P.M., Fisher, V., Foley, J., Friend, A.D., Kucharik, C., Lomas, M.R., Ramankutty, N., Sitch, S., Smith, B., White, A., Young-Molling, C., 2001. Global response of terrestrial ecosystem structure and function to CO₂ and climate change: results from six dynamic global vegetation models. *Global Change Biology* 7, 357–373.
- CSIR, 1999. South African National Land Cover Database Project. Data Users Manual (Final Report ENV/P/C 98136). CSIR, Pretoria, South Africa.
- Drake, B.G., Gonzalez-Meler, M.A., Long, S.P., 1997. More efficient plants: a consequence of rising atmospheric CO₂? *Annual Review of Plant Physiology and Plant Molecular Biology* 48, 609–639.
- Graham, R.W., Grimm, E.C., 1990. Effects of global climate change on the patterns of terrestrial biological communities. *Trends in Ecology and Evolution* 5, 289–292.
- Halpin, P.N., 1997. Global climate change and natural-area protection: management responses and research directions. *Ecological Applications* 7, 828–843.
- Hannah, L., Midgley, G.F., Lovejoy, T., Bond, W.J., Bush, M., Lovett, J.C., Scott, D., Woodward, F.I., 2002. Conservation of biodiversity in a changing climate. *Conservation Biology* 16, 264–268.
- Hastie, T.J., Tibshirani, R.J., 1990. *Generalized Additive Models*. Chapman & Hall, London.
- Hewitt, G., 2000. The genetic legacy of the quaternary ice ages. *Nature* 405, 907–912.
- Hughes, L., 2000. Biological consequences of global warming: is the signal already apparent? *Trends in Ecology and Evolution* 15, 56–61.
- Huntley, B., Birks, H.J.B., 1983. *An Atlas of Past and Present Pollen Maps for Europe, 0–13000 B.P.* Cambridge University Press, Cambridge.
- Le Maitre, D.C., Midgley, J.J., 1992. Plant reproductive ecology. In: Cowling, R.M. (Ed.), *The Ecology of Fynbos: Nutrients, Fire and Diversity*. Oxford University Press, Cape Town.
- Midgley, G.F., Roberts, R., 2001. Past climate change and the generation and persistence of species richness in a biodiversity hotspot, the Cape Flora of South Africa. In: Visconti, G., Beniston, M., Iannorelli, E.D., Barba, D. (Eds.), *Global Change and Protected Areas. Advances in Global Change Research*. Kluwer, Dordrecht, pp. 393–402.
- Midgley, G.F., Stock, W.D., Juritz, J.M., 1995. The effects of elevated CO₂ on Cape Fynbos species adapted to soils of different nutrient status: nutrient- and CO₂-responsiveness. *Journal of Biogeography* 22, 185–191.
- Midgley, G.F., Wand, S.J.E., Pammenter, N.W., 1999. Nutrient and genotypic constraints on CO₂-responsiveness: photosynthetic regulation in congeners of a nutrient-poor environment. *Journal of Experimental Botany* 50, 533–542.
- Midgley, G.F., Hannah, L., Roberts, R., MacDonald, D.J., Allsopp, J., 2001. Have Pleistocene climatic cycles influenced species richness patterns in the Greater Cape Mediterranean Region? *Journal of Mediterranean Ecology* 2, 137–144.
- Midgley, G.F., Hannah, L., Millar, D., Rutherford, M.C., Powrie, L.W. Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. *Global Ecology and Biogeography* (in press).
- Parkington, J., Cartwright, C., Cowling, R.M., Baxter, A., Meadows, M., 2000. Palaeovegetation at the last glacial maximum in the western Cape, South Africa: wood, charcoal and pollen evidence from Elands Bay Cave. *South African Journal of Science* 96, 543–546.
- Parmesan, C., 1996. Climate and species' range. *Nature* 382, 765–766.
- Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J.K., Thomas, C.D., Descimon, H., Huntley, B., Kaila, L., Kullberg, J., Tamaru, T., Tennet, W.J., Thomas, J.A., Warren, M., 1999. Poleward shifts in geographical ranges of butterfly species associated with global warming. *Nature* 399, 579–583.
- Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sanchez-Cordero, V., Soberon, J., Buddemeier, R.H., Stockwell, D.R., 2002. Future projections for Mexican faunas under global climate change scenarios. *Nature* 416, 626–629.
- Pimm, S.L., 2001. Entrepreneurial insects. *Nature* 411, 531–532.
- Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A., Solomon, A.M., 1992. A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography* 19, 117–134.

- Rebello, A.G., 2001. *Proteas*, a Field Guide to the Proteas of Southern Africa, second ed. Fernwood Press, Vlaeberg, Cape Town.
- Rouget, M., Richardson, D.M., Cowling, R.M., 2003. The current configuration of protected areas in the Cape Floristic Region—reservation bias and representation of biodiversity patterns and processes. *Biological Conservation* 112, 129–145.
- Rutherford, M.C., Westfall, R.H., 1994. Biomes of southern Africa—an objective categorization. *Memoirs of the Botanical Survey of South Africa* 63, 1–94.
- Schulze, R.E., 1997. *South African Atlas of Agrohydrology and Climatology (Water Research Commission Report TT82/96)*. Pretoria, South Africa.
- Schulze, R.E., Perks, L.A., 1999. *Assessment of the Impact of Climate. Final Report to the South African Country Studies Climate Change Programme*. School of Bioresources Engineering and Environmental Hydrology. University of Natal, Pietermaritzburg, South Africa.
- Shi, N., Dupont, L.M., Beug, H.J., Schneider, R.R., 1998. Vegetation and climate changes during the last 21,000 years in south-west Africa based on a marine pollen record. *Vegetation History and Archaeobotany* 7, 127–140.
- Stock, W.D., Midgley, G.F., 1995. Ecosystem response to elevated CO₂: nutrient availability and nutrient cycling. In: Moreno, J.M., Oechel, W.C. (Eds.), *Global Change and Mediterranean-type Ecosystems*. Ecological Studies 117. Springer-Verlag, New York, pp. 326–342.
- Taylor, H.C., 1978. Capensis. In: Werger, M.J.A. (Ed.), *Biogeography and Ecology of Southern Africa*. Junk, The Hague, pp. 171–229.
- Tyson, P., Odada, E., Schulze, R., Vogel, C., 2002. Regional–global change linkages: Southern Africa. In: Tyson, P., Fuchs, R., Fu, C., Lebel, L., Mitra, A.P., Odada, E., Perry, J., Steffen, W., Virja, H. (Eds.), *Global–Regional Linkages in the Earth System*. Springer, Berlin, pp. 3–73.
- Wand, S.J.E., Midgley, G.F., Millar, D. Recent temperature trends in the southwestern Cape, South Africa: potential implications for horticulture and natural vegetation. Personal communication.
- Warren, M.S., Hill, J.K., Thomas, J.A., Asher, J., Fox, R., Huntley, B., Roy, D.B., Telfer, M.G., Jeffcoate, S., Harding, P., Jeffcoate, P., Willis, S.G., Greatorex-Davies, J.N., Moss, D., Thomas, C.D., 2001. Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature* 414, 65–69.
- Yee, T.W., Mitchell, N.D., 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science* 2, 587–602.
- Younge, A., 2000. *Cape Action Plan for the Environment: A Biodiversity Strategy for the Cape Floral Kingdom*. World Wildlife Fund—South Africa, Cape Town, South Africa.
- Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms and aberrations in global climate 65 Ma to present. *Science* 292, 686–693.

ANNEXE 3.1

Niche properties and geographic extent as predictors of species sensitivity to climate change

Thuiller W., Lavorel S., Araújo M.B.

Soumis

Niche properties and geographic extent as predictors of species sensitivity to climate change

Wilfried Thuiller¹, Sandra Lavorel^{1,2} and Miguel B. Araújo¹

¹ Centre d'Ecologie Fonctionnelle et Evolutive. Centre National de Recherche Scientifique. 1919 route de Mende, 34293 Montpellier Cedex 5, France

² Laboratoire d'Ecologie Alpine, CNRS. Université J. Fournier. BP 53X, 38041 Grenoble Cedex 9, France

Correspondence author: Wilfried Thuiller
E-mail: wilfried.thuiller@cefe.cnrs-mop.fr

ABSTRACT

Common hypotheses on the effects of species ecology have so far not been considered by niche-based models to predict species sensitivity to climate change. Coupling bioclimatic models with an exploration of species niche characteristics and of their geographic exposure to climate change may provide an opportunity to deduce general and reproducible indicators of species sensitivity. We illustrate this for 1200 European plants species by relating the percentages of remaining/gained habitats by 2050 with their ecological niche properties (niche position, breadth and range size). Generally, niche position and breadth appeared as powerful indicators of species sensitivity, as they were linearly related to the percentage of remaining suitable habitat in the future. Range size and niche position were most suitable to

describe suitable habitats potentially gained by species. Finally relating species chorotypes to percentage of remaining/gained habitat provided insights into the biogeographic patterns of species sensitivity to climate change.

KEYWORDS

Bioclimatic modelling, Ecological niches, Climate change, Future potential habitats, Geographic exposure, Niche breadth, Niche position, Range size.

Running title: Predicting species sensitivity to climate change

Type of article: Report

Number of words in the abstract: 153

Number of words in the manuscript as a whole: 4187 (+ 6 figures/tables and 53 references)

INTRODUCTION

Climate change is already affecting biodiversity at various scales, for instance by increasing extinctions rates, by modifying phenology and physiology of species or by inducing displacements of species distributions (Hughes 2000; Parmesan & Yohe 2003). The exceptional rapidity of these changes makes it urgent to evaluate the potential effects of future climate change on biodiversity, a task best achieved using modelling tools. Since ten years, several studies have attempted to project species distributions into the future by using future climate change scenarios at large spatial scales (e.g. Huntley *et al.* 1995; Sykes 2001; Peterson *et al.* 2002). These studies made it possible to assess the sensitivity of species to climate change, defined by the degree to which species are affected negatively by climate change.

In spite of being broadly based on a multidimensional representation of ecological niches, species modelling tools have so far not been used to relate future potential species range shifts with species niche properties such as niche position and breadth. According to common ecological understanding, species that encounter a broader array of climatic conditions across their range are expected to have broader tolerances to climate change than restricted

species (Brown 1995). Likewise, species with requirements close to mean climatic conditions and with broad niche breadth should be less sensitive than species with outlying niches (marginal species) or with narrow niche breadth (Schoener 1989; Leibold 1995). Hence, niche properties and species' response to climate change may indeed be related, and niche properties could be used as indicators of sensitivity. Several hypotheses can be deduced from general ecological knowledge to predict sensitivity at large spatial scale: 1- Marginal species should be more sensitive to climate change than central ones (Swihart *et al.* 2003); 2- Species with narrow niche breadth should more sensitive to climate change than generalist species (Brown 1995); 3- Species with restricted ranges may be more sensitive to climate change than widespread species (Johnson *et al.* 1998).

In order to evaluate sensitivity of biodiversity to climate change, niche-based models are currently used to relate present-day species distributions with a set of climatic variables and project future distributions under climate change scenarios. However, while these models have produced estimates of species turnover, range expansion and contraction under climate change scenarios (Bakkenes *et al.* 2002; Peterson *et al.* 2002), no one has explored systematically the relationship

between niche properties of species (position, breadth, range size, chorotype) and their projected sensitivity to climate changes.

In this study, we used niche-based projections to examine relationships between species sensitivity and niche properties. We used the BIOMOD framework (Thuiller 2003) to project the future potential habitats of 1200 European higher plants according to one severe climate change scenario for 2050. Projected stable, disappearing and potentially gained climatically suitable habitats were then related to species niche characteristics and to species' geographic exposure to climate change. We addressed the following questions: (1) Are environmentally marginal species sensitive to climate change? (2) Are specialist species more sensitive than generalist to climate change? (3) Are species with restricted ranges more sensitive to climate change than widespread ones? (4) Are species from different chorotypes equally threatened?

Species data

Species presence-absence data were a subset (20%) of the Atlas Florae Europaeae (AFE) (Lahti & Lampinen 1999), which contains digitised maps for more than 2600 plants across Europe. For model accuracy, we removed under-represented species with less than 25 occurrences and thus used only 1200 species. Retained species included all European pteridophytes and a sample of spermatophytes comprising of all gymnospermae families (Coniferales, Taxales and Gnetales), a fraction of angiospermae dycotyledones (Salicales, Myricales, Juglandales, Fagales, Urticales, Proteales, Santales, Aristolochiales, Balanophorales, Polygonales, Centrospermae and Ranales), but no monocotyledones. Species data were originally located in 4419 UTM (Universal Transverse Mercator) 50x50 km grid cells, but we used only 2089 grid cells excluding most of the eastern European countries (except for the Baltic States) because of low recording efforts in these areas (Williams *et al.* 2000).

We defined 10 classes of chorology based on the phytogeographic and biogeographic properties of species: Alpine, Atlantic, Boreo-Alpine (Boreo-Subalpine and Artico-Alpine), Euro-Siberian (Medio-European), Introduced, strict Mediterranean, Mediterraneo-Atlantic (Mediterranean species having a wide part of their distribution in the Atlantic region), Late Mediterranean (Mediterranean species distributed significantly beyond the Mediterranean region), Coastal (species occurring exclusively on the European seaboard) and Pluriregional (species occurring in several biogeographic regions). The chorology of each species was determined according to the following references, completed and corrected by experts: Bolòs *et al.* (1990), ECOFLORA (<http://www.york.ac.uk/res/ecoflora/cfm/ecofl/index.cfm>), Flora Europaea (Tutin *et al.* 1964-1993) and CATMINAT (<http://perso.wanadoo.fr/philippe.julve/catminat.htm>).

Environmental data

We used seven climatically derived variables considered critical to plant physiological function and survival (Bartlein *et al.* 1986; Woodward 1987). Data were obtained from the Climatic Research Unit (<http://www.cru.uea.ac.uk/>) (Mitchell *et al.* unpublished) for: mean annual, winter and summer precipitation, mean annual temperature and minimum temperature of the coldest month, growing degree days (GDD>5°) and an index of humidity (AET/PET: mean ratio of annual actual over annual potential evapotranspiration). Mean values were averages for the period of 1961-1990.

Future projections for the 2050 time slice (averages for 2020-2050) were derived using one general circulation model experiment (GCM: Carson 1999) made available through the IPCC-DDC (Data Distribution Centre of the IPCC: ipcc-ddc.cru.uea.ac.uk) and conducted at the UK Hadley Centre for Climate Prediction and Research (HadCM3). This GCM projected mean climate under one rather extreme scenario (Nakicenovic & Swart 2000), the A1 storyline, which describes a future world of very rapid economic growth, global population that peaks by mid-century and then declines, and the rapid introduction of new and more efficient technologies. Under this scenario, atmospheric CO₂ should rise from 380 ppm in 2000 to 580 ppm in 2050. Mean annual temperature over Europe, which equals 8.3°C for the time period 1960-1990, is predicted to reach 10.4°C for 2020-2050.

Future species potential habitats

Future potential habitats were projected using the BIOMOD framework (Thuiller 2003). For each species, Generalised Linear Models (GLM), Generalised Additive Models (GAM), Classification Tree Analysis (CTA) and Artificial Neural Networks (ANN) were calibrated on a random sample of the initial data (70%). Then for every species the predictive accuracy of each model was assessed and compared using the Area Under the ROC Curve (AUC index) on the remaining 30% of the initial data (Pearce & Ferrier 2000; Thuiller *et al.* 2003a; Thuiller *et al.* 2003b). The best model for each species according to the ROC curve criterion was used to project future potential climatically suitable habitats. A rough guide for classifying the accuracy of model is the traditional academic point system: 0.90-1 = excellent; 0.80-0.90 = good; 0.70-0.80 = fair; 0.60-0.70 = poor; 0.50-0.60 = fail (Swets 1988).

For each species, we calculated the percentage of stable climatically suitable habitat defined as the grid cells suitable for the species both at present and under the scenario of change. Of the currently suitable envelope, the remaining grid cells, predicted to become unsuitable, made the percentage of lost habitat. Finally, the percentage of new climatically suitable habitat, defined as habitat unsuitable at present but predicted to become suitable, was calculated as the

ratio between the number of new climatically suitable grid cells and the present day habitat size. Potential new suitable habitat for a given species should not be interpreted as migration because we did not take into account species dispersal capacities.

Species niche properties and geographic exposure to climate change

The ecological niche of a species can be defined by its position and breadth along various environmental axes (Schoener 1989). Here, niche position and niche breadth along climatic axes were described using the multivariate co-inertia analysis Outlying Mean Index (OMI: Doledec *et al.* 2000). Unlike the more commonly used canonical correspondence analysis (ter Braak 1986) and redundancy analysis (Rao 1964), OMI makes no assumption about the shape of species response curves to the environment, and it gives equal weights to species-rich and species-poor sites. The result of this analysis is a description for each species of the mean niche position on each selected gradient (niche position), which represents a measure of the distance between the mean conditions used by the species and the mean conditions of the study area for that gradient. A species was then considered 'marginal' when its environment requirements are far from the mean European conditions. For every selected gradient, it is then possible to estimate the variability of habitat conditions used by each species by calculating the variance of the positions of each occurrence on the selected gradient (named niche breadth) (Thuiller *et al.* in press-c). Range size was described by the number of occurrences. Finally, the geographic exposure of each grid cell of the European map to climate change was estimated for each climatic variable using the coefficient of variation, calculated as: $100 * (\text{Var}_{\text{future}} - \text{Var}_{\text{current}}) / \text{mean}(\text{Var}_{\text{future}}, \text{Var}_{\text{current}})$.

To relate patterns of change across species with their niche properties, we ran stepwise generalised linear models using percentage of stable habitats and new potential habitats as response variables and niche properties as explanatory variables (niche position, niche breadth and range size). This procedure selected the most relevant variables among niche properties that could explain the overall sensitivity of a species to climate change according to the Akaike Information Criterion (Akaike 1974).

RESULTS

Bioclimatic gradients in European plant distributions

OMI extracted two main gradients explaining over 93% of species niche separation (Table 1). The first axis, explaining 80% of the total variation in species distributions, captured a temperature gradient running from the north-east to the south-west, parallel to the

Atlantic coast. The second gradient, explaining 13% of the variation, represented changes in seasonal and annual precipitation. For example the Mediterranean region had a high score for the first OMI axis (high temperature) and generally low scores for the second axis (low precipitation). Score variation along the second axis highlighted regional differences in moisture availability from dry (e.g. eastern coast of Spain) to wet (e.g. northwest of the Iberian Peninsula).

Future potential habitats

The predictive accuracy was high for almost all species, as AUC for the first quartile was higher than 0.9 and the mean AUC equal to 0.95. Models of current distributions could then be used to generate projections of future climatically suitable habitat. Consequences of climate change on the selected European plant species were predicted to be fairly severe (Table 2). On average, plants in this study would conserve only 70% of their current suitable habitat (30% of habitat loss). Extremes for stable habitats ranged from 0 (e.g. *Quercus fructicosa*) to 100 (e.g. *Dianthus fructicosus*) showing highly idiosyncratic responses across species. Some species could extend their current distributions by 30%, while others were predicted to gain no new suitable habitat (e.g. *Betula pubescens*) and other to expand by 400% (e.g. *Silene exaltata*).

Relationships between ecological niche properties and future projections

Stepwise GLM selected niche position on the temperature gradient (axis1), range size and niche breadth on the temperature gradient (axis 1) as the most explanatory variables for the percentage of stable climatically suitable habitat (Table 3; Figure 1A). The relationship between niche position on the temperature gradient and percentage of stable climatically suitable habitat was almost linear, showing that species occurring in colder regions (negative side) were expected to lose larger proportions of habitat than species from warmer regions (positive side). The bell-shaped relationship between range size and percentage of stable suitable habitat showed that species with very low numbers of occurrences were projected to lose large proportions of suitable habitat (45-60%), but that more abundant species were also expected to lose substantial amount of habitat (40-50%). Finally, the positive linear relationship between niche breadth on the temperature gradient and the remaining suitable habitat showed that species with narrow tolerance to temperature were expected to lose large proportions of habitats (40-60%) while species tolerant to aridity were predicted to be the most stable under future conditions.

Stepwise GLM selected range size, niche position on the temperature gradient (axis 1) and niche breadth on the precipitation gradient (axis 2) to explain variation in the percentage of potential habitats gained by species (Table 3; Figure 1B). Species with low

numbers of occurrences were expected to gain the largest proportion of potentially suitable habitat. The bell-shaped relationship between percentage of potentially gained habitat and niche position on the temperature gradient showed that marginal species (i.e. species from warmest or coldest habitats) were expected to gain the lowest proportion of new habitats, contrary to species living in temperate regions. The negative relationship for niche breadth on the precipitation gradient showed that to gain potential suitable habitat, species should have narrow niche breadth on this axis (i.e. species with narrow tolerance to precipitation were predicted to gain largest amount of new habitats). This relationship could result from the fact that Boreo-Alpine species have a wide precipitation tolerance while they were predicted to gain no potential habitats due to their marginal position on the temperature gradient.

Relationships between geographic exposure, species chorology and future projections

There was a strong heterogeneity in geographic exposure to climate change (Figure 2). Mean annual temperature, mean temperature of the coldest month and growing degree days (major component of the first OMI axis) were predicted to vary the most, especially in European mountain ranges and the Boreo-arctic region. The Mediterranean region was predicted to become drier, while precipitation in the Boreo-arctic region should remain stable.

Examining sensitivity among chorological groups suggested that, despite the large inter-specific variability within types, chorotypes susceptible to lose the largest amount of habitat were Atlantic, Alpine and Boreo-alpine (mean loss 55%), while Mediterranean species were projected to lose the lowest amount of habitat (Figure 3). Conversely, Mediterranean species were projected to potentially gain the highest proportion of habitat (median equal to 80%) while Boreo-Alpine species should gain the least (median equal to 8%).

This analysis of geographic exposure and species chorology helped us clarifying the observed relationships between niche properties and species sensitivity. For instance, marginal species in cold regions should be more sensitive to climate change while marginal species in warm regions would be less sensitive to temperature warming due to differing levels of exposure (Figure 1-A). This adds complexity to the original expectation that marginal species overall should be more sensitive than species using mean habitat conditions in Europe. Analysing the effect of exposure and chorology suggested that sensitivity depends strongly on the position of species niches on the climatic gradient undergoing change and the expected direction of the change. These two factors would for example explain why species tolerating warm temperatures were predicted to expand although they were marginal. These conclusions also apply to rare species, and to species with narrow niche breadth.

DISCUSSION

Methodological limitations

The BIOMOD and OMI analysis performed here depend critically on the modelled ecological niche as a stable constraint on the species' geographical distributional potential. The main assumption for modelling ecological niches is the equilibrium concept, which assumes that modelled species are at equilibrium with their environment (Prentice *et al.* 1992; Guisan & Theurillat 2000). Here we assumed that species were close to equilibrium at the spatial extent and resolution of our data set. In particular this was justifiable by the fact that, at the scale considered, impacts of land use/cover on species distributions modelling are overridden by climatic effects (Thuiller *et al.* in press-a).

Another limitation concerns the full description of the dimensions of species ecological niches. First, the analysis did not capture the full environmental range of some species to describe accurately their ecological niches (Pearson *et al.* 2002). This was especially the case for species at the edge of their climatic range, such as many eastern Mediterranean species. Second, projections into the future with climatic conditions outside of the current range that was used to calibrate models will necessarily involve uncertainties in the form of both over- and under-estimation (Thuiller *et al.* in press-b). Third, the degree to which interspecific interactions may limit current species distributions across large spatial scales can lead to under-estimating the potential distributions under current climate and hence disturb projections of future suitable habitats (Peterson 2003). Finally, the observed realised niche may exceed the fundamental niche of species (*sensu* Hutchinson 1957) if immigration from productive source areas to less suitable localisations is sufficiently large (Source-sink theory: Pulliam 2000). This bias would obviously disturb future projections, hence conclusions on species sensitivity to climate change.

Projections of potential suitable habitats are also strongly dependent on the resolution of the bioclimatic analysis. This might have been the case for Alpine species, as 50x50 km cells cannot reliably account for the heterogeneity and spatial structure of mountains, so that any conclusions on these species should be taken with care.

Finally, Thuiller (2003) demonstrated that different modelling techniques could lead to different projections into the future, depending on how each technique deals with the combination of environmental conditions to predict current and future distributions. The BIOMOD framework was proposed to address this problem by compiling four niche-based models for each species and selecting the best one based on a quantitative evaluation process (Thuiller 2003).

Exposure of European plant species to climate change

Global climate change was projected to strongly influence the future potential suitable habitats of the selected European higher plants. Even if our analysis

concerns only a small portion of total European plant diversity, we showed that the response of species to climate change can be informed by geographical and ecological considerations.

At the continental scale, exposure to climate change is geographically variable. Our analysis illustrated the basic idea that the sensitivity of a given species will depend on its geographical distribution and ecological properties relative to the regional patterns of exposure to climate change. A species with a niche position within regions most exposed to climate change will be bound to be more sensitive than a species in less exposed region. European climate was described by a dominant temperature gradient and a secondary precipitation gradient. Knowing the niche position of species on these gradients makes it possible to estimate potential pressures of climate change, especially for species with aggregated distributions or for those occurring in a single biogeographical region. In a warming context, species occurring preferentially in warm areas should benefit from climate change, conserving their initial habitats and/or expanding to new suitable habitats. Conversely, species occurring in cold regions should either disappear as a result of loss of suitable habitats, or migrate northwards or upwards towards new potential habitat.

Predicting species sensitivity from niche properties

Are marginal species sensitive to global climate change?

Our analysis demonstrated that marginal species are not always the most sensitive to global climate change as sensitivity is strongly related to exposure, niche breadth and range size. Marginal species like several Mediterranean species were even projected to lose little current suitable habitat and gain potential new suitable habitats. Our results concur with analyses for South Africa, in which most vulnerable species to climate change are either marginal species with disjointed ranges, or species occurring predominantly in western arid areas, likely to be the most exposed part of South-Africa (Erasmus *et al.* 2002). Marginality should always be related to the gradient considered, because species occupying specialised habitat are by definition more susceptible to changes in their habitats. For instance, Boreo-Alpine species are projected to lose important proportion of their current habitats and gain little new habitats. These species are highly marginal on the temperature gradient, which is the most susceptible to change, and have small niche breadths and low range size. They could therefore be considered as particularly sensitive to climate change. Likewise, for very marginal Alpine species with small niche breadth on the temperature gradient, BIOMOD projected a substantial loss of suitable habitats. The analysis, however, was not able to take into account spatial heterogeneity, aspect or slope, and should hence be interpreted with caution, as they are many potential topographic refuges and the possibly for species to migrate upslope (Gottfried *et al.* 1995; Peterson 2003).

Are specialist species more sensitive than generalists?

It was expected that species which encounter a broader array of climatic conditions across their range would have broader tolerances to climate change than restricted species (Brown 1995). On the temperature gradient, the most prone to change, our analysis corroborated this expectation. The linear relationship observed between the percentage of stable habitat and niche breadth showed that species occupying a large range of climatic conditions should be the most tolerant to climate change. However, these species were not predicted to gain large amounts of potential habitat because they could only expand their range within the limit of available habitats. As expected, species with narrow niche breadth were predicted to lose a high proportion of their current habitat if they occurred preferentially in cold and exposed parts of Europe (Swihart *et al.* 2003). This is mainly the case of Boreo-Alpine and some Euro-Siberian species. However, a species with narrow niche breadth, but with a niche position at the warmer end of the temperature gradient and living in a little exposed region should not be threatened by climate warming (e.g. Mediterranean species).

Are species with restricted ranges more sensitive to climate change?

Rare species occupy a central place in biodiversity concerns because they are the most prone to extinction (Pimm *et al.* 1988; Gaston 1994). We did not address particularly the risk of extinction of very narrowly distributed species as we removed the rarest species with fewer than 20 occurrences. However, for species with very restricted ranges, our projections provided evidence that climate change should contribute to increased extinction risk, as a result of a strong decrease in suitable habitat, especially for species most exposed to climate change. Nevertheless, species with restricted ranges were expected to potentially gain disproportionately substantial amounts of habitat when they lived in temperate regions. This curvilinear relationship between restricted ranges and percentage of new potential suitable habitat does however not take into account geographic distances and dispersal ability. Rarest species are often considered to have low dispersal ability, not to be strong competitors, and to have small local populations (Gaston 1994). All of these attributes would compromise the ability of species with restricted ranges to increase their range by colonising geographically disjunct new habitats.

CONCLUSION

This study highlighted general patterns about the relationships between sensitivity of species to climate change and their ecological properties. These patterns appear most strongly driven by the exposure of species to climate change, with additional effects of species niche characteristics. We were able to draw the following generalisations:

- 1) Niches are powerful indicators of species' sensitivity to climate change
- 2) Boreo-Alpine species, being marginal at the cold end of the temperature gradient, having narrow niche breadth and often restricted ranges are predicted to be highly sensitive as they occur in the most exposed regions to climate warning.
- 3) Alpine species, which are also marginal at the cold end of the temperature gradient, but have larger niche breadth than Boreo-Alpine, are under pressure by climate change but could also gain large amount of suitable habitat by upwards migration, a feature that could not be captured at the resolution of our models.
- 4) Mediterranean species, which are at the warm end of temperature gradient, with medium niche breadth and medium range size, were predicted to lose proportionally less suitable habitat and gain substantial amount of new habitats. Even if these species do not have wide niches, they are not in very exposed regions to climate warning.

ACKNOWLEDGMENTS

This research was funded by the European Commission's ATEAM (Advanced Terrestrial Ecosystem Analysis and Modelling) project (EVK2-CT-2000-00075). We thank T.D. Mitchell of the Tyndall Centre for Climate Change Research for providing climate and scenario data, and S. Zaehle and M. Erhard at the Potsdam Institute for Climate Impact Research for aggregating climate data. Lluís Brotons and Sébastien Lavergne provided valuable comments throughout the preparation of the manuscript.

CITED REFERENCES

Akaike H. (1974) A new look at statistical model identification. *IEEE Transactions on Automatic Control*, AU-19, 716-722

Bakkenes M., Alkemade R.M., Ihle F., Leemans R. & Latour J.B. (2002) Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Glob. Change. Biol.*, 8, 390-407

Bartlein P.J., Prentice I.C. & Webb T. (1986) Climatic response surfaces from pollen data for some eastern North American taxa. *J. Biogeogr.*, 13, 35-57

Bolòs O.d., Vigo J., Masalles R.M. & Ninot J.M. (1990) *Flora manual dels Països Catalans*. Pòrtic, Barcelona.

Brown J.H. (1995) *Macroecology*. The University of Chicago Press, Chicago.

Carson D.J. (1999) Climate Modelling: achievements and prospects. *Q.J.R. Meteorol. Soc.*, 125, 1:27

Doledec S., Chessel D. & Gimaret-Carpentier C. (2000) Niche separation in community analysis: a new method. *Ecology*, 81, 2914-2927

Erasmus B.F.N., Van Jaarsveld A.S., Chown S.L., Kshatriya M. & Wessels K.J. (2002) Vulnerability of South African animal taxa to climate change. *Glob. Change. Biol.*, 8, 679-693

Gaston K.J. (1994) *Rarity*. Chapman and Hall, London.

Gottfried M., Pauli H., Reiter K. & Grabherr G. (1995) A fine-scaled model for changes in species distribution patterns of high mountain plants induced by climate warming. *Divers. Distrib.*, 5, 241-251

Guisan A. & Theurillat J.-P. (2000) Equilibrium modelling of alpine plant distribution: how far can we go? *Phytocoenologia*, 30, 353-384

Hughes L. (2000) Biological consequences of global warming: is the signal already apparent? *TREE*, 15, 56-61

Huntley B., Berry P.M., Cramer W. & McDonald A.P. (1995) Modelling present and potential future ranges of some European higher plants using climate response. *J. Biogeogr.*, 22, 967-1001

Hutchinson G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 145-159

Johnson D.D.P., Hay S.I. & Rogers D.J. (1998) Contemporary environmental correlates of endemic areas derived from meteorological satellite sensors. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.*, B 265, 951-959

Lahti T. & Lampinen R. (1999) From dot maps to bitmaps - Atlas Florae Europaeae goes digital. *Acta Bot. Fennica*, 162, 5-9

Leibold M.A. (1995) The niche concept revisited: mechanistic models and community context. *Ecology*, 76, 1371-1382

Nakicenovic N. & Swart R. (2000) *Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge.

Parnesan C. & Yohe G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37-42

Pearce J. & Ferrier S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.*, 133, 225-245

Pearson R.G., Dawson T.P., Berry P.M. & Harrison P.A. (2002) SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecol. Model.*, 154, 289-300

Peterson A.T. (2003) Projected climate change effects on Rocky Mountain and Great Plain birds: generalities of biodiversity consequences. *Glob. Change. Biol.*, 9, 647-655

Peterson A.T., Ortega-Huerta M.A., Bartley J., Sanchez-Cordero V., Soberon J., Buddemeier R.H. & Stockwell D.R.B. (2002) Future projections for Mexican faunas under global climate change scenarios. *Nature*, 416, 626-629

Pimm S.L., Jones H.L. & Diamond J. (1988) On the risk of extinction. *Am. Nat.*, 132, 757-785

Prentice I.C., Cramer W., Harrison S.P., Leemans R., Monserud R.A. & Solomon A.M. (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.*, 19, 117-134

Pulliam H.R. (2000) On the relationship between niche and distribution. *Ecol. Lett.*, 3, 349-361

- Rao C.R. (1964) The use and interpretation of principal correspondence analysis in applied research. *Sankhya A*, 26, 329-359
- Schoener T.W. (1989) The ecological niche. In: *Ecological Concepts* (ed. Cherret, J.). Blackwell, pp. 790-813.
- Swets K.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293
- Swihart R.K., Gehring T.M. & Kolozsvary M.B. (2003) Responses of 'resistant' vertebrates to habitat loss and fragmentation: the importance of niche breadth and range boundary. *Divers. Distrib.*, 9, 1-18
- Sykes M.T. (2001) Modelling the potential distribution and community dynamics of lodgepole pine (*Pinus contorta* Dougl. ex . Loud.) in Scandinavia. *For. Ecol. Manage.*, 141, 69-84
- ter Braak C.J.F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67, 1167-1179
- Thuiller W. (2003) BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change. *Glob. Change. Biol.*, 9, 1353-1362
- Thuiller W., Araújo M.B. & Lavorel S. (2003a) Generalized Models versus Classification Tree Analysis: a comparative study for predicting spatial distributions of plant species at different scales. *J. Veg. Sci.*, 14, 669-680
- Thuiller W., Araújo M.B. & Lavorel S. (in press-a) Do we need land-cover data to model species distributions in Europe? *J. Biogeogr.*
- Thuiller W., Brotons L., Araújo M.B. & Lavorel S. (in press-b) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*
- Thuiller W., Lavorel S., Midgley G.F., Lavergne S. & Rebelo A.G. (in press-c) Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* species in the Cape Floristic Region. *Ecology*
- Thuiller W., Vaydera J., Pino J., Sabaté S., Lavorel S. & Gracia C. (2003b) Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecol. Biogeogr.*, 12, 313-325
- Tutin T.G., Burges N.A., Chater A.O., Edmondson J.R., Heywood V.H., Moore D.M., Valentine D.H., Walters S.M. & Webb D.A. (1964-1993) *Flora Europaea*. Cambridge University Press, Cambridge.
- Williams P., Humphries C., Araújo M.B., Lampinen R., Hagemeyer W., Gasc J.-P. & Mitchell-Jones T. (2000) Endemism and important areas for representing European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belg. J. Entomol.*, 2, 21-46
- Woodward F.I. (1987) *Climate and plant distribution*. Cambridge University Press, Cambridge.

Table 1. Canonical correlation between climatic variables and the two first axes from the OMI.

Climatic variables	OMI axis 1	OMI axis 2
Growing degree day	0.97	0.02
Annual temperature	0.94	-0.16
Mean Temp. of the Coldest month	0.87	-0.37
Index of humidity	- 0.86	-0.23
Summer precipitation	- 0.74	-0.60
Annual precipitation	- 0.21	- 0.97
Winter precipitation	0.15	- 0.92

Table 2. Summary of BIOMOD projections for 1200 European higher plants in 2050 under the A1FI HadCM3 future scenario

	% stable suitable habitat	% of potentially new suitable habitat
Min	0%	0%
1 st quartile	53.3%	13.1%
Median	70.7%	29.5%
3 rd quartile	86.6%	59.4%
Max	100%	414.1%

Table 3. Summary of stepwise GLM carried out on % of stable and gained habitats. Df : degree of freedom; resid : residual; AIC : Akaike Information Criteria (Akaike 1974).

Step	Df	Deviance resid.	Df resid.	Deviance	AIC
<i>% of stable habitats</i>					
Poly (Mean position on T° gradient, 2)	2	38.5	1185	250	256
Poly(Range size, 2)	2	25.9	1183	224	234
Niche breadth on T° gradient	1	2.5	1182	222	234
<i>% of gaining habitats</i>					
Poly(Range size, 2)	2	15238	1185	36642	36648
Poly(Mean position on T° gradient, 2)	2	8703	1183	27938	27948
Poly(Niche breadth on Pr. Gradient, 2)	2	3428	1181	24511	24525

Figure 1. Response curves from stepwise GLM relating % of stable (a) and potentially new suitable habitat (b) with significant species ecological niche properties, plotted in a decrease of explained deviance order.

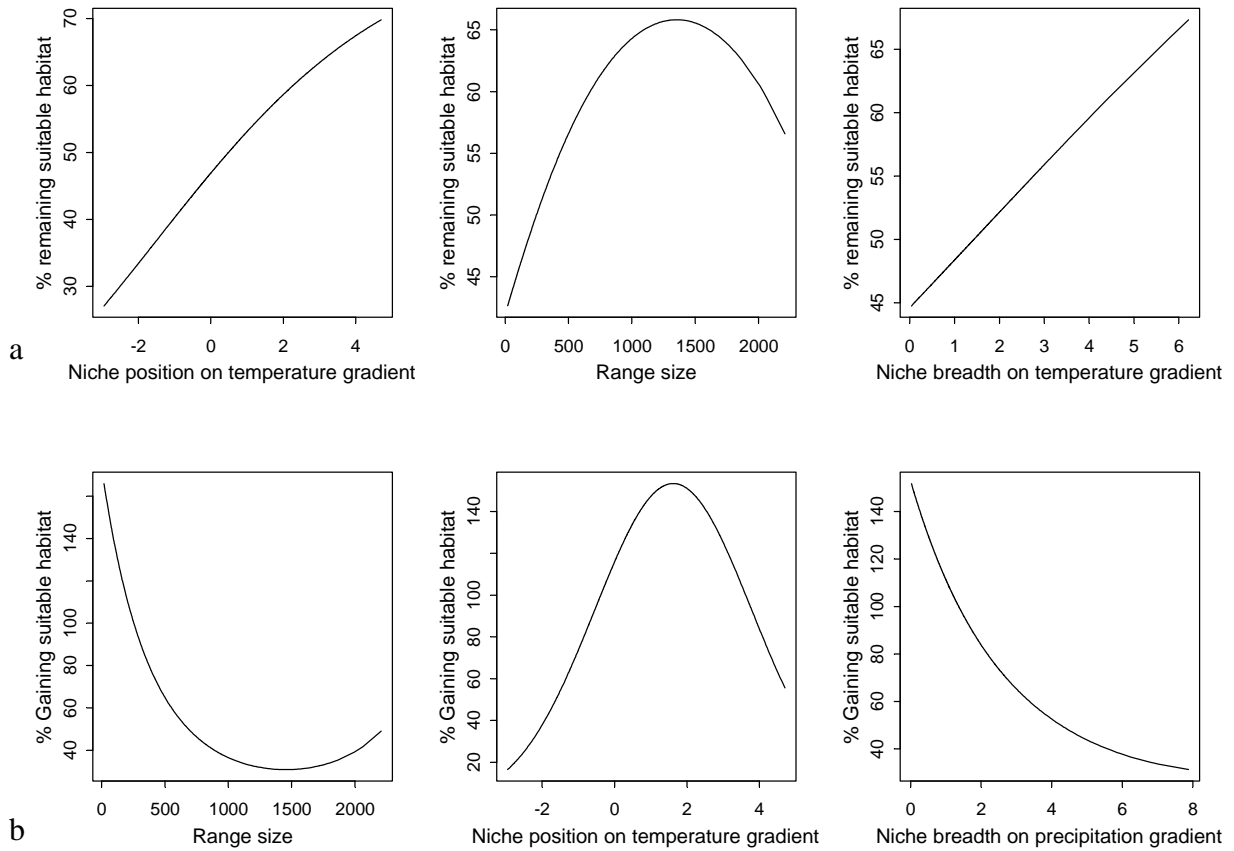
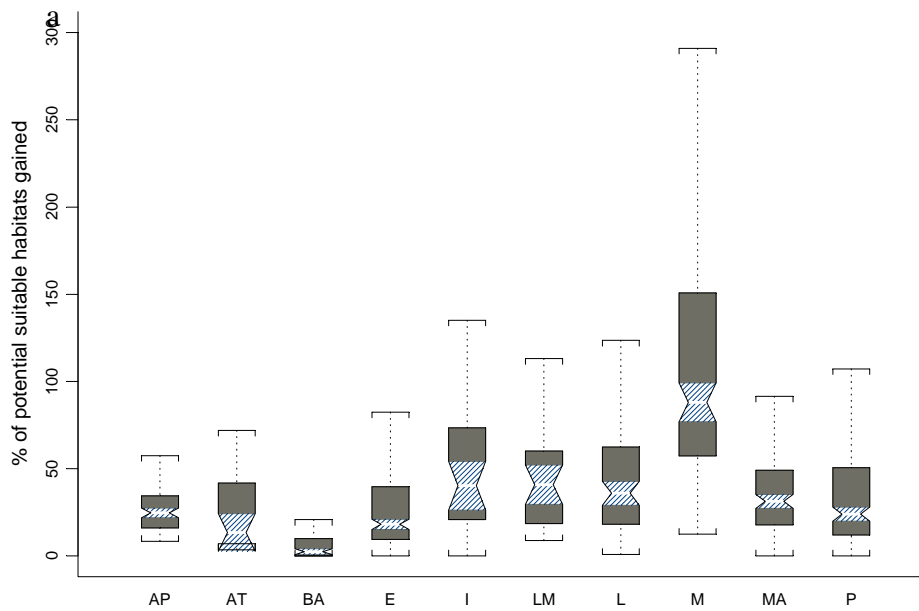
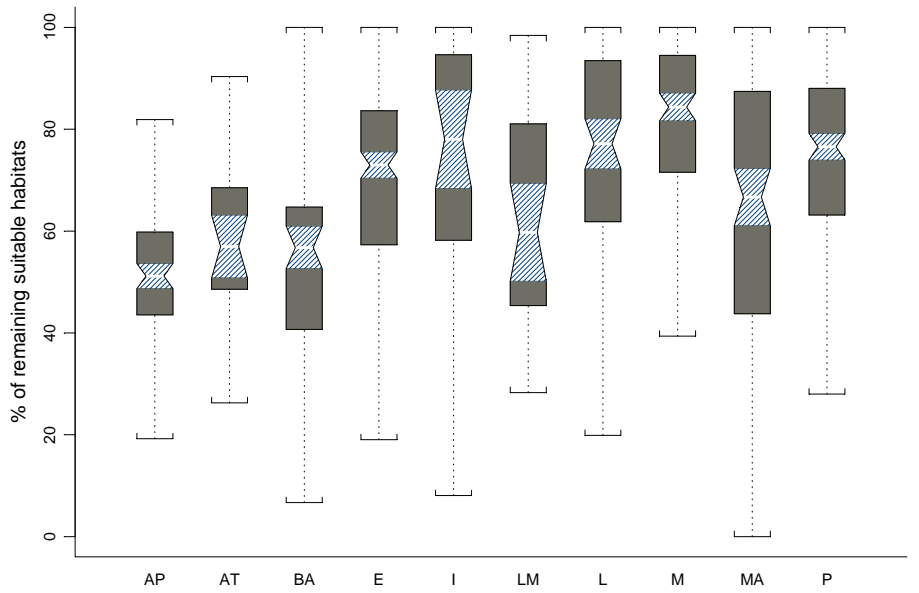


Figure 2: Geographic exposure of regions to climate change for every climatic variable.
 AET/PET: ratio of actual on potential evapotranspiration



Figure 3. Relationship between species chorotypes and (a) percentage of stable suitable habitats and (b) percentage of new potential suitable habitats gained by species. AP: Alpine; AT: Atlantic; BA: Boreo-Alpine; E: Euro-Siberian; I: Introduced; LM: Late-Mediterranean; L: Littoral; M: Mediterranean; MA: Mediterraneo-Atlantic; P: Pluri-regional species. Hatched shapes correspond to confidence intervals at 95%.



b

ANNEXE 3.2

Uncertainties in species range shifts projections under global climate change scenarios

Thuiller W., Araújo M.B., Brotons L., Lavorel S., Midgley G.F., Williams P.H.

En préparation

Uncertainties in species range shifts projections under global climate change scenarios

Wilfried Thuiller*, Miguel B. Araújo*, Lluís Brotons*, Sandra Lavorel**†, Guy F. Midgley§ and Paul Williams‡

* CEFE-CNRS, 1919 route de Mende, 34293 Montpellier cedex 5, † Laboratoire d'Ecologie Alpine, CNRS, Université J. Fournier, BP 53X, 38041 Grenoble cedex 9, France, § Center for Applied Biodiversity Science, Conservation International, 1919 M St., NW, Washington, D.C. 20036, USA, ‡Biogeography and Conservation Laboratory, The Natural History Museum, London, UK.

Global climate change is threatening many species with extinction⁽¹⁻³⁾. Recent insight into consequences on biodiversity has been gained from niche-based models that describe the potential responses of individual species to climate change by relating their present-day distribution to current climate, and projecting spatial shifts in their climatic envelopes according to climate change scenarios good⁽⁴⁻⁶⁾. Here we show that different types of niche-based models produce different results, adding unexpected and previously not measured uncertainties to projections⁽⁶⁾. Using five climate change scenarios and four different niche-based models, combined with two alternative

methodological assumptions, we project the future climatic envelopes (for 2050) of 2013 European plant and animal species. All models suggest high potential species turnover (20-40%) in response to climate change. However variability in the distributional change of species across climate scenarios is obscured by a strong variability in species projections arising from alternative, yet equally justifiable, niche-based models. Therefore, projections of future species distributions and derived community descriptors cannot be reliably discussed in an ecological context unless model uncertainty is explicitly measured.

Significant effects of global climate change are already observed for a variety of ecosystems, especially with evidence for shifts in species ranges^(7,8). Because different species respond idiosyncratically to various ecological forces, ecological communities may disassemble as individual species shift their ranges in different directions. To assess the potential impact of climate change on communities, the percentage of species turnover (defined as an index of dissimilarity between the current and future species communities within a given area) is often considered as a good measure of the degree of ecosystem perturbation, as well as the percentage of species which persist in, disappear from, and colonise that area⁽⁴⁾. Using climate change scenarios and niche-based models that project future suitable habitat from current distributions, several studies have suggested that species turnover may be very high in some regions (about 40%), resulting in strong modifications of community structure and possibly leading to ecosystem disruption^(4,5,9). However, absolute turnover rates (e.g. at continental scales) are difficult to interpret and act upon if their spatial pattern is not known. For management purposes, projections of species turnover should be reliable and highlight precisely where strong modifications of ecosystem will be expected. This requires that spatially explicit information on the distribution and accuracy of estimated turnover rates be generated at finer (e.g. regional) scales. Sources of uncertainty may include

climate change scenarios and different methodologies to obtain maps of turnover rates.

In order to quantify projection uncertainties that may arise from using different input parameters or techniques, we generated species climatic envelopes (for 2050) for 2013 species (383 birds, 162 mammals, 118 herptiles, 1305 plants) in Europe, under present conditions and five future climate change scenarios (ranging from rapid global economic and population growth to low regional economic and population growth scenarios). We used four familiar niche-based modelling techniques, combined with two robust methods to transform probabilities of occurrence generated by these models into binary presence/absence data (cut-off method). The factorial combination of methodological approaches and climate scenarios yielded 40 future climatic envelopes for every species. Combinations of individual species envelopes for a given method and scenario allowed us to project the associated species turnover. The analysis of these three sources of variability in the model outputs allows us to estimate the mean expected species turnover from the models as well as the spatial structure of confidence in model predictions

The mean across projections of species turnover for individual pixels ranges from 20% to 41%, with the absolute minimum from 5% to 21% and the absolute maximum from 44% to 72%. Although, species

turnover is high across Europe it shows high variability across the 40 projections, implying that delivering a single mean value for one area offers very little information about the turnover rate to be expected. Furthermore, the value of mean species turnover over the 40 projections is highly spatially heterogeneous (Figure 1). Mediterranean regions (Spain, south-eastern France, northern Italy and the Balkan Peninsula) are consistently predicted to be the most sensitive to global climate change. Mean species turnover is also high in the Fenno-scandinavian region, while the biodiversity of the Euro-Siberian region should not be strongly perturbed by climate change. Some localities in southeastern France and Italy have a mean species turnover in excess of 50%. Nevertheless there is some doubts on the reliability of mean turnover value across the 40 projections. An absolute estimate of uncertainty across the 40 projections can be obtained by the coefficient of variation of species turnover (Figure 2-A). This uncertainty is maximal in some areas of high species turnover (Spain, France, Fenno-Scandinavia), suggesting that conclusions about species turnover and impacts of climate change on ecosystems may be highly contextual.

When partitioning the variation in the coefficient of variation according to the different sources of uncertainty, it is the type of niche-model that shows the highest contribution (Figure 2-B-C-D), with a high correlation to total uncertainty (Spearman rank correlation $\rho = 0.79$, $p < 0.001$). Moreover these maps highlight a large uncertainty resulting from each of the three sources, making general statements on the sensitivity of the biodiversity of any given area to climate change largely invalid. Will Mediterranean region experience high (~40%) or moderate species turnover (~20%)? To what extent would Alpine faunas and floras be replaced by lower altitude species? Several such crucial questions for the future of European biodiversity are not straightforward to answer when spatial uncertainty is high.

Such estimate would be equal to obtain a spatially consistent turnover pattern across all the models use. We generated such “consensus analysis” by using the first factor of the principal correspondence analysis (PCA) of the 40 projections across all pixels to summarise common patterns. ANOVA on the first PCA axis values confirms that the niche-based modelling method explains the largest part of measured uncertainty (deviations from an hypothetical consensus estimate in which all models would be equivalent), followed by the cut-off method and finally climate change scenarios (Table 1). Furthermore, the fact that interaction terms are all significant indicates that the detected effects of climate change scenarios on spatial patterns of species turnover depend on the methodological approach to project climatic envelopes. There is no consistent pattern across different types of models for a given climate change scenario and vice versa.

In our approach, the first PCA axis can also be used as a reliable conservative projection of spatial patterns of species turnover across Europe including only consistent information derived from the different methodologies and climate scenarios used. As such it could provide a conservative estimate of expected community changes that may serve as a spatially explicit base for conservation and other management decisions on (Figure 3). For instance, this consensus across projections shows that the northwest of the Iberian Peninsula, southeastern France, northwestern Italy, Yugoslavia and northern Finland are expected to undergo the strongest changes in species composition, while the biodiversity of Germany, Poland, Denmark and the Netherlands is expected to change least. Nevertheless, as only 20% of the total European flora was used and observed distributions of animal data are not totally reliable, such projections are restricted to the species included in the analysis and are only an illustration of what it could happen in short term future.

Modelling future changes in species distribution and community structure is and will be limited by uncertainty originating both from methodological approaches and the different socioeconomic hypotheses used to derive future climate change scenarios. In this context, our objective should be to integrate rather than ignore such uncertainty, and to generate consensus estimates that represent consistent patterns across models and scenarios and are hence robust. The only way to achieve this objective is to explicitly include all significant sources of uncertainty in modelling frameworks and to derive conservative predictions of spatial patterns of species turnover. Previous attempts to quantify changes in community structure at large spatial scales are likely to be seriously flawed since they did not explicitly account for important uncertainty sources in the modelling exercise^(4,5,9).

1. Chapin III, S. F. et al. Consequences of changing biodiversity. *Nature* **405**, 234-242 (2000).
2. Sala, O. E. et al. Global biodiversity scenarios for the year 2100. *Science* **287**, 1770-1774 (2000).
3. Walther, G.-R. et al. Ecological responses to recent climate change. *Nature* **416**, 389-395 (2002).
4. Peterson, A. T. et al. Future projections for Mexican faunas under global climate change scenarios. *Nature* **416**, 626-629 (2002).
5. Bakkenes, M., Alkemade, R. M., Ihle, F., Leemans, R. & Latour, J. B. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Glob. Change. Biol* **8**, 390-407 (2002).
6. Thuiller, W. BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change. *Glob. Change. Biol* **00**, 000-000 (2003).

7. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37-42 (2003).
8. Root, T. L. et al. Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57-60 (2003).
9. Erasmus, B. F. N., Van Jaarsveld, A. S., Chown, S. L., Kshatriya, M. & Wessels, K. J. Vulnerability of South African animal taxa to climate change. *Glob. Change. Biol* **8**, 679-693 (2002).
10. Jalas, J. & Suominen, J. *Atlas Florae Europaeae* (The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki, 1972-1996).
11. Mitchell-Jones, A. J. et al. *Atlas of European Mammals* (Academic Press, London, 1999).
12. Gasc, J.-P. et al. *Atlas of amphibians and reptiles in Europe* (Societas Europaea Herpetologica & Museum National d'Histoire Naturelle, Paris, 1997).
13. Hagemeyer, W. J. M. & Blair, M. J. *The EBCC atlas of European breeding birds, their distribution and abundance* (Poyser, London, 1997).
14. Araújo, M. B. et al. Would environmental diversity be a good surrogate for species diversity? *Ecography* **24**, 103-110 (2001).
15. Humphries, C. et al. Plant diversity in Europe: *Atlas Florae Europaeae* and WORLDMAP. *Acta Bot. Fennica* **162**, 11-21 (1999).
16. Lahti, T. & Lampinen, R. From dot maps to bitmaps - Atlas Florae Europaeae goes digital. *Acta Bot. Fennica* **162**, 5-9 (1999).
17. Williams, P. et al. Endemism and important areas for representing European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Bel. J. Ent.* **2**, 21-46 (2000).
18. Mitchell, T. D. A comprehensive set of climate scenarios for Europe. *In prep* (2002).
19. Carson, D. J. Climate Modelling: achievements and prospects. *Q.J.R. Meteorol. Soc.* **125**, 1:27 (1999).
20. Nakicenovic, N. & Swart, R. (eds.) *Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, 2000).
21. Hill, J. K., Thomas, C. D. & Huntley, B. Climate and habitat availability determine 20th century changes in a butterfly's range margin. *Proc. R. Soc. Lon. B* **266**, 1197-1206 (1999).
22. Frescino, T. S., Edwards Jr., T. C. & Moisen, G. G. Modeling spatially explicit forest structural attributes using Generalized Additive Models. *J. Veg. Sci.* **12**, 15-26 (2001).
23. Leathwick, J. R. Climatic relationships of some New Zealand forest tree species. *J. Veg. Sci.* **6**, 237-248 (1995).
24. Rouget, M. et al. Determinants of distribution of six *Pinus* species in Catalonia, Spain. *J. Veg. Sci.* **12**, 491-502 (2001).
25. Thuiller, W. et al. Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecol. Biogeogr.* **12**, 313-325 (2003).
26. Moisen, G. G. & Frescino, T. S. Comparing five modelling techniques for predicting forest characteristics. *Ecol. Model.* (In press).
27. Pearson, R. G., Dawson, T. P., Berry, P. M. & Harrison, P. A. SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecol. Model.* **154**, 289-300 (2002).
28. Peterson, A. T. Projected climate change effects on Rocky Mountain and Great Plain birds: generalities of biodiversity consequences. *Glob. Change. Biol* **9**, 647-655 (2003).
29. Fielding, A. H. & Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Env. Conserv.* **24**, 38-49 (1997).
30. Monserud, R. A. & Leemans, R. Comparing global vegetation maps with the Kappa statistic. *Ecol. Model.* **62**, 275-293 (1992).

Table 1. Summary of the ANOVA carried out on the first axis of the consensus analysis. Df represents the degree of freedom, Sum of sq, the sum of squares and Mean sq, the mean square of the ANOVA. F value corresponds to the mean square of the component divided by the mean square of residual. Probability values quantify the significance of the tests.

	Df	Sum of sq	Mean Sq	F value	Pr(f)
Model	3	0.1146	0.0382	655.5	< 0.0001
Threshold	1	0.0285	0.0285	489.4	< 0.0001
Scenario	4	0.0827	0.0207	354.6	< 0.0001
Model x Threshold	3	0.0116	0.0039	66.2	< 0.0001
Model x Scenario	12	0.0093	0.0077	13.3	< 0.0001
Threshold x Scenario	4	0.0047	0.0012	20.3	< 0.0001
Residuals	12	0.0007	0.0001		

Figure 1. Modelled mean percentage of species turnover across the 40 projections. Red colours highlight regions with highest projected turnover, while blue colours correspond with the lowest projected turnover.

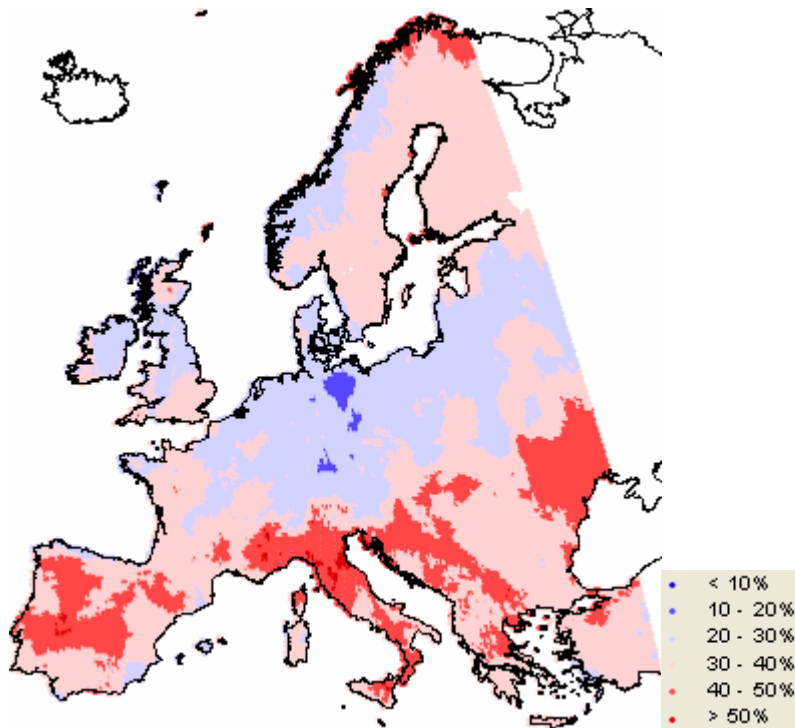


Figure 2. Uncertainty of mean species turnover represented by the coefficient of variation ($100 \times \text{standard deviation} / \text{mean}$) across all 40 projections (%). A) Total uncertainty; B) Uncertainty across the four niche-based models; C) Uncertainty across the two cut-off methods; D) Uncertainty across the five global change scenarios.

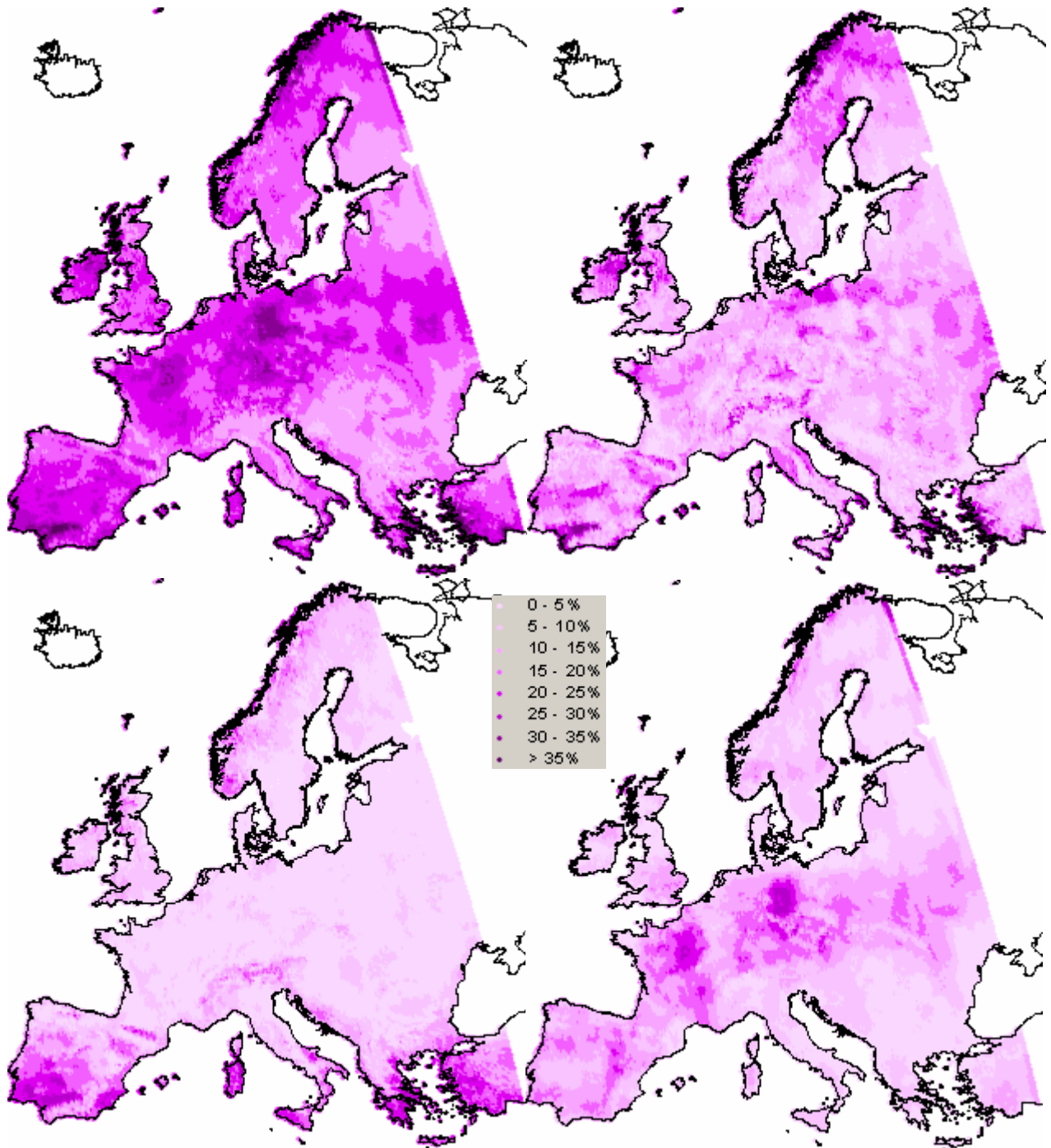
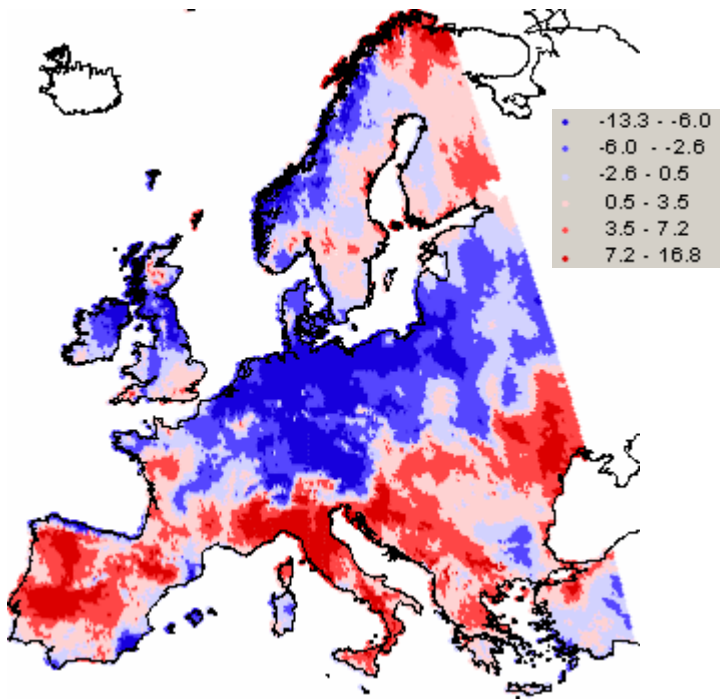


Figure 3. Map of the values of the first axis of the PCA across all 40 projections. Dark red colours correspond to high rates of species turnover; blue colours correspond to lower species turnover rates.



METHODS

Data on current species distributions:

Original species data include 2,294 plant⁽¹⁰⁾, 186 mammal⁽¹¹⁾, 143 amphibian and reptile⁽¹²⁾ and 440 breeding bird species⁽¹³⁾. Terrestrial vertebrates include all known species⁽¹⁴⁾ whereas plants comprise only ~20% of the European flora⁽¹⁵⁾. The grid used was based on the *Atlas Florae Europaeae*⁽¹⁶⁾, with cells boundaries typically following the 50 km lines of the Universal Transverse Mercator (UTM) grid. The mapped area comprises western, northern and southern Europe, but excludes most of the eastern European countries where recording effort was both less uniform and less intensive⁽¹⁷⁾. After removing species with less than twenty records necessary for model accuracy, the data set comprised 1305 plants, 162 mammals, 118 herptiles and 383 birds.

Current and future climate data:

Climate data used for analyses were a comprehensive set of bioclimatic variables for Europe from the Climatic Research Unit (<http://www.cru.uea.ac.uk/>)⁽¹⁸⁾ (<http://www.pik-potsdam.de/ateam/>): mean annual, winter and summer precipitation, mean annual temperature and minimum temperature of the coldest month, growing degree days (>5°) and an index of humidity (mean ratio of annual actual over annual potential evapotranspiration). Means were averaged for the period of 1961-1990. These data sets were supplied on a 10' grid, covering Europe. Then data were aggregated by averaging to 50x50 km UTM in order to match the resolution of species data. Niche-based models were calibrated on the 50x50 km UTM grid and then species distributions projected on the 10' grid for current and future climate.

Future projections were derived using two different general circulation model for period 2050⁽¹⁹⁾: The GCM experiments (HadCM3 and CSIRO-Mk2) make available through the IPCC-DDC (Data Distribution Centre of the IPCC: ipcc-ddc.cru.uea.ac.uk) and they are respectively conducted at the UK Hadley Centre for Climate Prediction and Research (HadCM3) and the Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO). These GCM included different storylines scenarios according to IPCC (Intergovernmental Panel on Climate Change)' recommendations⁽²⁰⁾: The A1 storyline describes a future world of very rapid economic growth, global population that peaks in mid-century and decline after, and the rapid introduction of new and more efficient technologies. The A2 storyline describes a very heterogeneous world. The underlying theme is self-reliance and preservation of local identities. Economic development is primarily regionally oriented and per capita economic growth and technological change are more fragmented and slower than the other storylines. The B1 storyline describes a convergent world with the same global population that peaks in mid-century and declines thereafter, as the A1 storyline, but with rapid change in

economic structures toward a service and information economy, with the introduction of clean and resource-efficient technology. The B2 storyline describes a world in which the emphasis is on local solutions to economic, social, and environment sustainability. It is a world with continuously increasing global population at a rate lower than A2, intermediate levels of economic development, and less rapid and more diverse technological change than in the B1 and A1 storylines⁽²⁰⁾.

Niche-based models of species climatic envelopes:

Niche-based models relate the present-day distribution of species to a pool of available environmental variables using statistical or rule-based methods. Different modelling techniques are commonly applied to project potential future species distributions. Although Generalised Linear Models (GLM) are the most common^(5,21), Generalised Additive Models (GAM) are being used increasingly^(22,23). Alternative rule-based approaches have also emerged as interesting tools to predict current and future potential distributions. These include Classification And Regression Tree analysis (CART)^(24,25), Artificial Neural Networks (ANN)^(26,27) or Genetic Algorithms^(4,28).

We used BIOMOD, a framework which capitalizes on the different techniques (GLM, GAM, ANN, CTA) used for niche-based modelling to provide alternative spatial predictions⁽⁶⁾. We simulated for each species current and future distributions for all models and used both the Relative Operating Characteristic Curve⁽²⁹⁾ and the Kappa statistic⁽³⁰⁾ to select an optimal threshold to transform continuous probability values from models into presence-absence.

Percentage of species turnover:

Each species was predicted to lose or gain suitable habitat, depending on the model, cut-off method and climate change scenario. To evaluate geographic patterns of numbers of species lost or gained across Europe, we summed the number of species gained (G) or lost (L) by pixel across all models, methods and scenarios. We summed simulated present-day distributions across all species to determine the simulated species richness (SR) and estimated the likely percentage of species turnover as: $100(G+L)/(SR+G)$.

Consensus analysis:

This was based on a principal component analysis (PCA) ran on the 40 model x threshold x climate scenario combinations of species turnover (or percentage of species gained or lost by species). The first axis of PCA analysis allowed to capture consistent spatial patterns in turnover rate across the different projections and thus to remove idiosyncratic variability from projections' disagreement. If the explanatory power of the first axis was strong (~100% of explained variance), we could argue that projections are spatially consistent with a first axis loadings very correlated to

En préparation

mean species turnover of all singular projections. Inversely, if the explained variance of the first axis was low (<50% of explained variance), we could conclude that model projections largely disagree. This disagreement could be visualized analysing the following axes of the PCA. In the latter case, it is unlikely that the first axis correlates to the general mean due to the variability across projections.

If projections are reliable, we expect the first axis to be strongly correlated to the general mean. We analysed factors inducing potential disagreement between projections (model methodology, threshold choice, and climate scenario) by means of ANOVA on the loadings of the first PCA axis.

ANNEXE 3.3

Patterns and uncertainties of species' range shifts under global change

Thuiller W.

Soumis

Patterns and uncertainties of species' range shifts under climate change

Wilfried Thuiller

CEFE-CNRS, 1919 route de Mende, 34293 Montpellier cedex 5, France

Corresponding author: wilfried.thuiller@cefe.cnrs-mop.fr

ABSTRACT

Although bioclimatic modelling is often used to estimate potential impacts of likely climate changes, little has been done to assess the reliability and variability of projections. Here, using four niche-based models, two methods to derive probability values from models into presence-absence data and five climate change scenarios, I project the future potential habitats of 1305 European plant species (for 2050). All 40 different projections of species turnover across Europe suggested high potential species turnover (20-40%) in response to climate change. However variability in the potential distributional changes of species across climate scenarios was obscured by a strong variability

in projections arising from alternative, yet equally justifiable, niche-based models. Therefore, projections of future species distributions and derived community descriptors cannot be reliably discussed unless model uncertainty is quantified explicitly. I propose and test an alternative way to account for modelling variability to derive estimates of species turnover (with and without dispersal) according to a range of climate change scenarios representing various socio-economic futures..

Keywords: BIOMOD, climate change scenarios, consensus analysis, niche-based models, plant diversity, species turnover, uncertainty.

Introduction

Significant effects of global climate change are already observed for a variety of ecosystems, especially with evidence for shifts in species ranges (Parmesan & Yohe 2003; Root *et al.* 2003). Because different species respond idiosyncratically to various ecological forces, ecological communities may disassemble as individual species shift their ranges in different directions. Recent insights into consequences of climate change on biodiversity have been gained from niche-based models that describe the potential responses of individual species to climate change by relating their present-day distribution to current climate, and projecting spatial shifts in their climatic envelopes according to climate change scenarios (Bakkenes *et al.* 2002; Peterson *et al.* 2002; Thuiller 2003). The percentage of species turnover (defined as an index of dissimilarity between the current and future species composition within a given area), as well as the percentage of species which could persist in, disappear from, and colonise that area are often considered as a good measures of the degree of ecosystem perturbation, and have been already used to assess the potential impact of climate change at large spatial scales (Peterson *et al.* 2002). Using climate change scenarios and niche-based models that project future suitable habitat from current distributions, several studies have suggested that species turnover may be very high in some regions (about 40%), resulting in strong modifications of community structure that may lead to ecosystem disruption (Bakkenes *et al.* 2002; Erasmus *et al.* 2002; Peterson *et al.* 2002).

However, projections of species turnover should be reliable and robust to be able to highlight precisely where strong modifications of species diversity are expected to occur. Although niche-based models have been extensively assessed in their effectiveness to predict present-day distributions of various taxa from regional to continental scales (Guisan *et al.* 1999; Moisen & Frescino 2002; Thuiller *et al.* 2003), their accuracy in predicting future conditions has not been evaluated yet. Due to the obvious reasons that realised future species distributions are unknown at present, reliability of niche-based models for projecting species distributions to future hypothetical conditions should be not taken for granted (Thuiller *et al.* 2004). Recently, Thuiller (2003) showed that a range of niche-based models with similar accuracy for predicting species' present-day distribution could predict substantially different future distributions. Such results have critical implications when results for individual species are used together to derive turnover or extinction rates as used by Peterson *et al.* (2002) or Bakkenes *et al.* (2002). Indeed, if for a given species, projections from two models into the future differ to a great extent, chances are high that this variability for deriving species turnover are exacerbated by using large number of species. Hence, there is an urgent need to assess the impact of variability across modelling methods used to derive conclusions about the impact of global change on biodiversity. This need increases when projections into the future are carried out according to a range of climate change scenarios as provided by the last report of the IPCC (Nakicenovic & Swart 2000).

Here, using five climate change scenarios and four different niche-based models, I project the future climatic envelopes (for 2050) of 1305 European plant species. Combinations of individual species envelopes for a given method and scenario allowed us to project the associated species turnover and percentage of stable/gained habitats. The analysis of these different sources of variability in the model outputs allows us to estimate separately the proportion of variability introduced by methodological issues and by climate change scenarios. Then, I propose a method to incorporate methodological uncertainty into the modelling process and derive robust conservative estimates of species turnover across a range of climatic scenarios.

Methods

The general approach to modelling climate change effects on species distributions used herein is developed in detail elsewhere (Thuiller 2003).

Plant species datasets

Species presence-absence data were a subset (20%) of the Atlas Florae Europaeae (AFE) (Lahti & Lampinen 1999), which contains digitised maps for more than 2600 plants across Europe. For model accuracy, I removed under-represented species with less than 20 occurrences and thus used only 1305 plant species. Retained species included all European pteridophytes and a sample of spermatophytes comprising of all gymnospermae families (Coniferales, Taxales and Gnetales), a fraction of angiospermae dycotyledones (Salicales, Myricales, Juglandales, Fagales, Urticales, Proteales, Santales, Aristolochiales, Balanophorales, Polygonales, Centrospermae and Ranales), but not monocotyledones. Species data were located in 4419 UTM (Universal Transverse Mercator) 50x50 km grid cells. I used only 2089 grid cells excluding most of the eastern European countries (except for the Baltic States) because of low recording efforts in these areas (Williams *et al.* 2000).

Bioclimatic datasets

I used seven climatically derived variables considered critical to plant physiological function and survival (Bartlein *et al.* 1986; Woodward 1987). Data were obtained from the Climatic Research Unit (<http://www.cru.uea.ac.uk/>) for: mean annual, winter and summer precipitation, mean annual temperature and minimum temperature of the coldest month, growing degree days (GDD>5°) and an index of humidity (AET/PET: mean ratio of annual actual over annual potential evapotranspiration). Mean values were averages for the period of 1961-1990. These data sets are supplied on a 10' grid, covering Europe. Then data were aggregated by averaging to 50x50 km UTM in order to match the resolution of species data.

Climate changes scenarios

Future projections were derived using two different general circulation model for period 2050 (Carson

1999): The GCM experiments (HadCM3 and CSIRO2) make available through the IPCC-DDC (Data Distribution Centre of the IPCC: ipcc-ddc.cru.uea.ac.uk) and they are respectively conducted at the UK Hadley Centre for Climate Prediction and Research (HadCM3) and the Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO2). These GCM included different storylines according to IPCC (Intergovernmental Panel on Climate Change)' recommendations (Nakicenovic & Swart 2000): The A1 storyline describes a future world of very rapid economic growth, global population that peaks in mid-century and decline after, and the rapid introduction of new and more efficient technologies. The A2 storyline describes a very heterogeneous world. The underlying theme is self-reliance and preservation of local identities. Economic development is primarily regionally oriented and *per capita* economic growth and technological change are more fragmented and slower than in the other storylines. The B1 storyline describes a convergent world with the same global population as the A1 storyline that peaks in mid-century and declines thereafter, but with rapid change in economic structures towards a service and information economy, with the introduction of clean and resource-efficient technology. The B2 storyline describes a world in which the emphasis is on local solutions to economic, social, and environment sustainability. It is a world with continuously increasing global population at a rate lower than A2, intermediate levels of economic development, and less rapid and more diverse technological change than in the B1 and A1 storylines (Nakicenovic & Swart 2000). I used the four storylines under HadCM3 GCM experiments and only A2 for CSIRO2.

Niche based-models

Current at 50x50 km and current and future potential habitats at 10° resolution were projected for each plant species using the BIOMOD framework (Thuiller 2003). For each species, Generalised Linear Models (GLM), Generalised Additive Models (GAM), Classification Tree Analysis (CTA) and Artificial Neural Networks (ANN) were calibrated on a random sample of the initial data (70%) and tested on the remaining datasets with both Receiver Operating Characteristic curve and *Kappa* statistic (Pearce & Ferrier 2000; Thuiller *et al.* 2003).

Presence-absence projected distributions were derived from probability values from niche-based models using two methodological approaches (cut-off methods hereafter): first, by maximising the *Kappa* statistic, and second by maximising the percentage of presence and absence correctly predicted for present conditions (Manel *et al.* 2001).

Each species was predicted to lose or gain suitable habitat according to each combination of modelling and cut-off methods on one hand and climate change scenarios on the other. To evaluate geographic patterns of numbers of species potentially lost or gained across

Europe, I summed the number of species gained (G) or lost (L) by pixel for each model. I summed simulated present-day distributions across all species to determine the simulated species richness (SR) and estimated the percentage of species turnover as: $100(G+L)/(SR+G)$. As this process implied unlikely universal dispersal, I also derived species turnover without dispersal as $100(L/SR)$. That means that species will not be able to follow suitable bioclimatic range shifts.

Quantifying uncertainties

The three factors tested here (4 niche-based models x 2 cut-off methods x 5 climate change scenarios) resulted in 40 projections of species turnover. To assess and account for the differences in sources of variability amongst projections, I used a consensus analysis (Westerhuis *et al.* 1998). This is based on a multivariate approach in order to search among a set of variables the one that summarises the highest amount of information. I used a principal component analysis (PCA consensus) on the different combinations of modelling method scenario to run a consensus analysis on projections of species turnover. The first axis (called consensus axis hereafter) of PCA captured consistent spatial patterns in turnover rate across the different projections. If the 40 projections were exactly similar, i.e. only one axis would be retained explaining 100% of the variation, the variability was null across projections. If the 40 projections were randomly simulated and completely unrelated to each other, 40 axes would be retained, each of them explaining $1/40 = 2.5\%$ of the variation. Based on the basic expectation that at least climate change scenarios were different, the first axis should be comprised between 2.5 and 100%. If the consensus axis explained a variability smaller than 100%, then the following axes (variability axes hereafter) would explain the orthogonal variation due to variability across the models, cut-off methods or/and scenarios of climate change. I analysed these variability axes using classification tree analysis (CTA, Breiman *et al.* 1984) to measure the proportion of variability explained by each of factor. I used CTA because this method is not perturbed by interrelated components and make it possible to separate variation within each factor (for instance between scenario A1 and scenario A2).

Results

Quantifying uncertainties

The consensus PCA was carried out on the 40 estimations of species turnover. The first five axes explained 90% of the total variability across the projections (Table 1). The consensus axis accounted for 56.1% of variability, i.e. the 40 projections were 56.1% similar. Hence, the variability across the three factors was equals to 43.9 % and described by the variability axes of the PCA.

The position of each projection and associated sources of variation (models x cut-off method x scenarios) along each of the four variability axes was extracted

and used to run CTA (Figure 1). The second, third and fourth orthogonal axes were essentially explained by the model and cut-off method used, while the fifth axis which account only for 4% of the variability was mainly driven by variability across scenarios. The second axis (Figure 1-A), which explained the greatest part of variability across turnover projections (18.3%), was mainly driven by differences between rule-based and generalised regression methods. The PCA 2nd axis scores, plotted on the European map, showed that this variability across these methods was spatially explicit. High scores (darker colours) showed high species turnover predicted by generalised models (GAM and GLM) and low scores (lighter colours) showed high species turnover predicted by rule-based models (ANN and CTA) but not by the generalised models. The third and fourth axes which explained noticeably less variability were also mainly determined by variability across Models and then by Cut-off methods.

Variability across climate change scenarios was discernable only for the fifth axis separating broadly the GCM (CSIRO2 *versus* HadCM3) and then the storylines (global *versus* regional economy). The PCA 5th axis scores, plotted on the European map, displayed higher species turnover predicted under HadCM3 than under CSIRO2.

Integrating modelling uncertainties

As I showed previously, modelling uncertainty was too high to derive reliable conclusion about species turnover across a range of climate change scenarios. To decrease this variability I used the PCA consensus but carried out for the couple Model x Cut-Off method for each selected species (8 projections per species) according to the range of climatic change scenarios (HadCM3 A1-A2-B1-B2 and CSIRO2 A2). The projections from the couple Model x Cut-Off method the most correlated with the PCA consensus axis represents the projections that best captures shared information amongst the whole set of projections. To derive species turnover rate across the five climate change scenarios, I selected for each species the couple Model x Cut-Off method the most correlated with the first PCA axis.

Percentage of species turnover in each pixel assuming no or universal dispersal ability showed interesting contrasts across the range of climate change scenarios (Figure 2 & 3). The main difference between projections was related to the GCM employed. Under the A2 storyline and especially in France, CSIRO2 had lower species turnover rates by pixel than HadCM3 (Fig. 2-3 B-C).

There were consistent patterns across the four storylines using HadCM3. While the regions where high rate of species turnover were predicted to be approximately the same, the intensity and extent were different under the A1 storyline, often the most critical for species distributions (Table 2). For instance, the biogeographical region the most affected by climate change in terms of species turnover with no or universal species dispersal ability were the Pannonian

regions under all storylines. Across all GCMs and storylines, rates of species turnover were noticeably lower for western and northern than southern and eastern biogeographic regions, showing that there was a front located at the borderline between Eurosiberian - Atlantic - and Mediterranean regions.

Under the more likely no dispersal scenario, Boreo-alpine region should be the least affected by climate change with a lower rate of species turnover than the other scenarios. Conversely, under universal dispersal, this region could be prone to high species turnover with high rate of colonisation by pixel. Under universal scenario, this was the Atlantic region the least affected by climate change under almost all scenarios.

While the observed general trend that storyline A1 (global and economic future with the highest CO₂ increase) was more threatening to plant diversity matched expectations (Nakicenovic & Swart 2000), it was less obvious that B2 and B1 storylines had a higher impact on plant diversity than the A2 storyline. For Alpine, Arctic and Boreal regions, the B2 storyline caused the highest rate of species turnover. Conversely, the A2 storyline (HadCM3 and CSIRO2) was projected as the scenario in which the lower number of changes in plant diversity were expected to occur.

Discussion

Modelling uncertainty

I do not discuss in this paper about the modelling limitations as recent papers provided insightful comments, descriptions and critics of such modelling approach (see Peterson *et al.* 2002; Pearson & Dawson 2003; Thuiller 2003).

In the present paper, I analysed the impact of variability across niche-based models and the impact of cut-off method used to derive probability values from niche-based models into presence-absence form. Although cut-off methods are largely under-studied, their impact on quality of projections appeared crucial as both methods (maximising Kappa statistic and maximising % of presence and absence correctly predicted) could lead to very different projected distributions, a result that could be problematic for biodiversity risk assessment and conservation purpose (Manel *et al.* 2001).

Here I demonstrated that even when using common niche-based modelling and cut-off methods, the variability across projections were huge and may hide the variability of using varying climate change scenarios as recommended by IPCC (Nakicenovic & Swart 2000). My aim here is not to be pessimistic on the suitability for niche-based models to derive broad estimates of sensitivity of biodiversity to climate change but to argue that care must be applied before deriving any conclusion about future impacts on species diversity. Recently, several authors published interesting results on species sensitivity to climate change and derived geographic values of species extirpation, colonisation or turnover (Bakkenes *et al.*

2002; Erasmus *et al.* 2002; Peterson *et al.* 2002). For instance using GLM, the same niche-based models as Bakkenes *et al.* (2002), I found that on average, the percentage of plants that were present in a pixel in 1990 and would disappear from that pixel by 2050 up to 41%. This rate is higher than estimations from Bakkenes *et al.* (2002) for the same flora (32%) although they include under-sampled area in Russia (Williams *et al.* 2000). However, the same estimator but using artificial neural networks was about 29%. Moreover, differences in estimated rates of species extinction or colonisation across niche-based models are spatially structured showing that models do not have the same ability to deal with some particular species or geographic locations with specific environmental conditions.

Since, there is no universal niche-based model better than all others (Franklin 1998; Thuiller *et al.* 2003), one could use for each species the model (across a panel of niche-based models) with the highest predictive accuracy for present-day conditions and project future distributions according to a range of climate change scenarios. However, there are at least two problems: 1- there are several methods to assess predictive accuracy (e.g. ROC curve or Kappa statistic) and they do often not provide the same estimates as they are based on different algorithms and assumptions (Manel *et al.* 2001). 2- it is not certain that the model with the highest predictive accuracy for present-day conditions will be the best to estimate future distributions (Thuiller 2003). Therefore, it could be safer to choose a projection (one niche-based model x one cut-off selection) into the future which conservatively summarises agreements amongst projections generated by different modelling techniques. The projection the most correlated with the first PCA axis represents such a consensus by extracting shared variations amongst future projections. It does not mean that it will necessarily be the most accurate projection into the future but we may argue that this is the most consensual and conservative one to generate robust hypothesis about future changes in species distribution.

Integrating modelling uncertainties

Using the above methodology, I demonstrated that species extirpation, colonisation and turnover could be derived according to a range climate change scenarios. Although climate change scenarios are based on multiples inputs, they attempted to account for economic, demographic and societal response to climate change (Nakicenovic & Swart 2000). They are useful for testing the varying response of plant species. I showed that the highest variability across response comes from the use of different GCM scenarios (CSIRO2 and HadCM3). The projections from CSIRO2 are much more conservative for plants diversity than from HadCM3. This is an important point as it demonstrates that there is an inherent variability across different GCMs based on the uncertainty in forecasts of anthropogenic climate

change (Allen *et al.* 2000). This uncertainty is then reflected on the response of species to climate change. Such known variability across GCMs strengthens the basic idea of reducing the uncertainty across species projections due to species modelling tools for forecasting impact of climate change on species diversity.

Across emission scenarios describing the societal response to climate change (Nakicenovic & Swart 2000), the A1 storyline appeared as the most threatening plant species diversity and enhanced species turnover across some specific regions such as southeast France or northern Italia. The A2 appeared less harmful for plant diversity than B1 and B2 representing a more environmentally-minded world. Indeed, at human time scale, climate is changing continuously and incidence of present pollution will be effective at least one or two decades later. That means that in average from 2020 to 2050, there is less CO₂ and NO₂ emissions under A2 than B1 and B2 storylines but that the trend is reverse from 2050 to 2080 (Nakicenovic & Swart 2000). That trend explains why A2 could be considered as safer for species diversity than B1 and B2 scenarios, although it is considered as one of the most negative scenarios for global equilibrium (Nakicenovic & Swart 2000).

It is often stressed that under climate change, northward and upward migration of species is expected (Parmesan & Yohe 2003) but little attention has been given to which regions would be the most affected by high species turnover and community changes. Under the presented simulations it appears that Pannonian and Continental biogeographical regions could undergo dramatic rate of species turnover (with and without dispersal) which could strongly modify existing species community and probably create new ones. However, results highlight that species dispersal could strongly moderate future species turnover. The simulations presented here show the range of possible between no dispersal and universal dispersal ability. I demonstrate that conclusions could be seriously different according to the dispersal scenario used, as Boreo-Alpine region could be the least perturbed assuming no dispersal, but one of the most affected assuming universal dispersal. Such phenomenon is expected in region that could undergo massive immigrations from neighbouring regions as Boreo-Alpine or Pannonian regions.

To conclude, I demonstrate that the methodological uncertainty associated to the niche-based models is problematic for forecasting impact of global climate change on biodiversity. This uncertainty hides the effects of using different GCMs and different emissions scenarios as recommended by IPCC (Watson & *et.* 2001). There is an inherent uncertainty for forecasting anthropogenic climate change and so predictions of current and future species distributions should be exempt of uncertainty in order to produce relevant patterns of extinctions, colonisations and species turnover. In this paper, I propose an alternative

way to incorporate species modelling uncertainty in order to produce more reliable estimates of species turnover, with and without species dispersal, under global climate change.

Acknowledgements: This research was funded by the European Commission's ATEAM (Advanced Terrestrial Ecosystem Analysis and Modelling) project (EVK2-CT-2000-00075). Species data were kindly supplied by M.B. Araújo. The author thanks particularly T.D. Mitchell of the Tyndall Centre for Climate Change Research for providing climate and scenario data and S. Zaehle and M. Erhard of Potsdam-Institute for Climate Impact Research for aggregating climate

CITED REFERENCES

- Allen MR, Stott PA, Mitchell JFB *et al.* (2000) Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature* **407**, 617-620.
- Bakkenes M, Alkemade RM, Ihle F *et al.* (2002) Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology* **8**, 390-407.
- Bartlein PJ, Prentice IC, Webb T (1986) Climatic response surfaces from pollen data for some eastern North American taxa. *Journal of Biogeography* **13**, 35-57.
- Breiman L, Friedman JH, Olshen RA *et al.* (1984) *Classification and regression trees*. Chapman and Hall, New York.
- Carson DJ (1999) Climate Modelling: achievements and prospects. *Q.J.R. Meteorol. Soc.* **125**, 1:27.
- Erasmus BFN, Van Jaarsweld AS, Chown SL *et al.* (2002) Vulnerability of South African animal taxa to climate change. *Global Change Biology* **8**, 679-693.
- 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.
- Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modelling of plant species distribution. *Plant Ecology* **143**, 107-122.
- Lahti T, Lampinen R (1999) From dot maps to bitmaps - Atlas Florae Europaeae goes digital. *Acta Botanica Fennica* **162**, 5-9.
- Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* **38**, 921-931.
- Moisen GG, Frescino TS (2002) Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling* **157**, 209-225.
- Nakicenovic N, Swart R, editors. (2000) Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on

Climate Change. Cambridge University Press, Cambridge.

- Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37-42.
- Pearce J, Ferrier S (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling* **128**, 127-147.
- Pearson RG, Dawson TE (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology & Biogeography* **12**, 361-372.
- Peterson AT, Ortega-Huerta MA, Bartley J *et al.* (2002) Future projections for Mexican faunas under global climate change scenarios. *Nature* **416**, 626-629.
- Root TL, Price JT, Hall KR *et al.* (2003) Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57-60.
- Thuiller W (2003) BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* **9**, 1353-1362.
- Thuiller W, Araújo MB, Lavorel S (2003) Generalized Models versus Classification Tree Analysis: a comparative study for predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* **14**, 669-680.
- Thuiller W, Brotons L, Araújo MB *et al.* (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography* **in press**.
- Watson RT, *et. al*, editors. (2001) IPCC Third assessment report: Climate change 2001. Synthesis Report. IPCC, Geneva, Switzerland.
- Westerhuis JA, Kourti T, MaxGregor JF (1998) Analysis of multiblock and hierarchical pca and pls methods. *Journal of Chemometrics* **12**, 463-482.
- Williams P, Humphries C, Araújo MB *et al.* (2000) Endemism and important areas for representing European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belgian Journal of entomology* **2**, 21-46.
- Woodward FI (1987) *Climate and plant distribution*. Cambridge University Press, Cambridge.

Table 1 Summary of the PCA consensus carried out on the forty projections.

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Variance	4.747	2.704	1.637	1.424	1.241
Proportion of Variance	0.561	0.183	0.067	0.057	0.039
Cumulative Proportion x 100	56.1	74.4	81.1	86.2	90.0

Table 2 Mean rate of species extirpation (Ext), colonisation (Col) and turnover (Tur) by pixel across biogeographical regions of Europe according to the range of climate change scenarios.

Région biogéographiqu e	HadCM3 A1			Csiro2 A2			HadCM3 A2			HadCM3 B1			HadCM3 B2		
	Ext	Col	Tur	Ext	Col	Tur	Ext	Col	Tur	Ext	Col	Tur	Ext	Col	Tur
Alpine	16	33. 8	36. 3	12. 9	25. 7	30. 1	14. 1	32	34	15. 5	32	34. 8	17. 4	33. 4	37. 1
Artic	14. 8	44. 3	39. 4	12. 5	26. 8	30. 3	15. 1	40. 2	37. 9	16. 4	44. 7	40. 4	18. 4	41. 4	40. 8
Atlantic	17. 1	27. 9	34. 3	12. 3	21	27. 3	14. 7	23	30. 2	16	24	31. 6	16. 8	24. 2	32. 5
Boreal	14. 5	53. 2	43. 8	14. 5	48. 6	41. 7	13. 3	49. 2	41. 6	13. 2	52. 9	42. 8	16. 2	51	44. 1
Continental	25. 1	36. 9	44. 9	16. 4	27. 1	33. 5	20. 7	31. 7	39. 1	21. 4	29. 7	38. 9	23. 3	31. 1	40. 9
Mediterranean	18. 7	25. 4	34. 8	15. 4	20. 3	29. 2	17. 3	23. 5	32. 6	19. 8	22. 1	33. 9	19. 1	23. 9	34. 4
Pannonian	27. 3	47. 1	50. 5	19. 8	46. 8	45. 1	22. 5	49	47. 9	27. 6	37. 2	47. 2	26. 8	47. 6	50. 3
Steppic	29. 5	45. 1	51. 4	26. 5	48. 9	50. 4	28	52. 9	52. 9	29. 8	39. 5	49. 7	29. 6	48. 9	52. 7

Figure captions

Fig 1. Classification Tree Analysis carried out on the position of each triplet (niche-based model x cut-off method x scenarios) on PCA axes 2, 3, 4 and 5 and the mapped scores of each axes.

Fig 2. Rate of plant species turnover in Europe assuming no species dispersal across a range of climate change scenarios. A) A1 HadCM3; B) A2 CSIRO2; C) A2 HadCM3; D) B1 HadCM3; E) B2 HadCM3

Fig 3. Rate of plant species turnover in Europe assuming universal dispersal across a range of climate change scenarios. A) A1 HadCM3; B) A2 CSIRO2; C) A2 HadCM3; D) B1 HadCM3; E) B2 HadCM3

Fig. 1

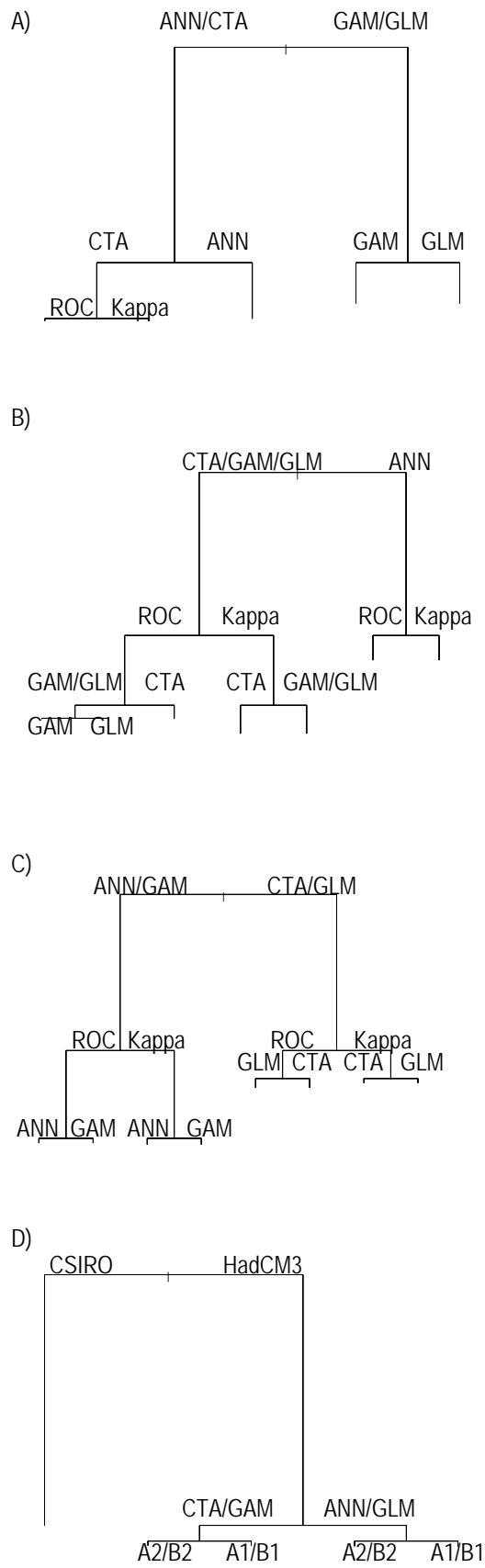


Fig. 2

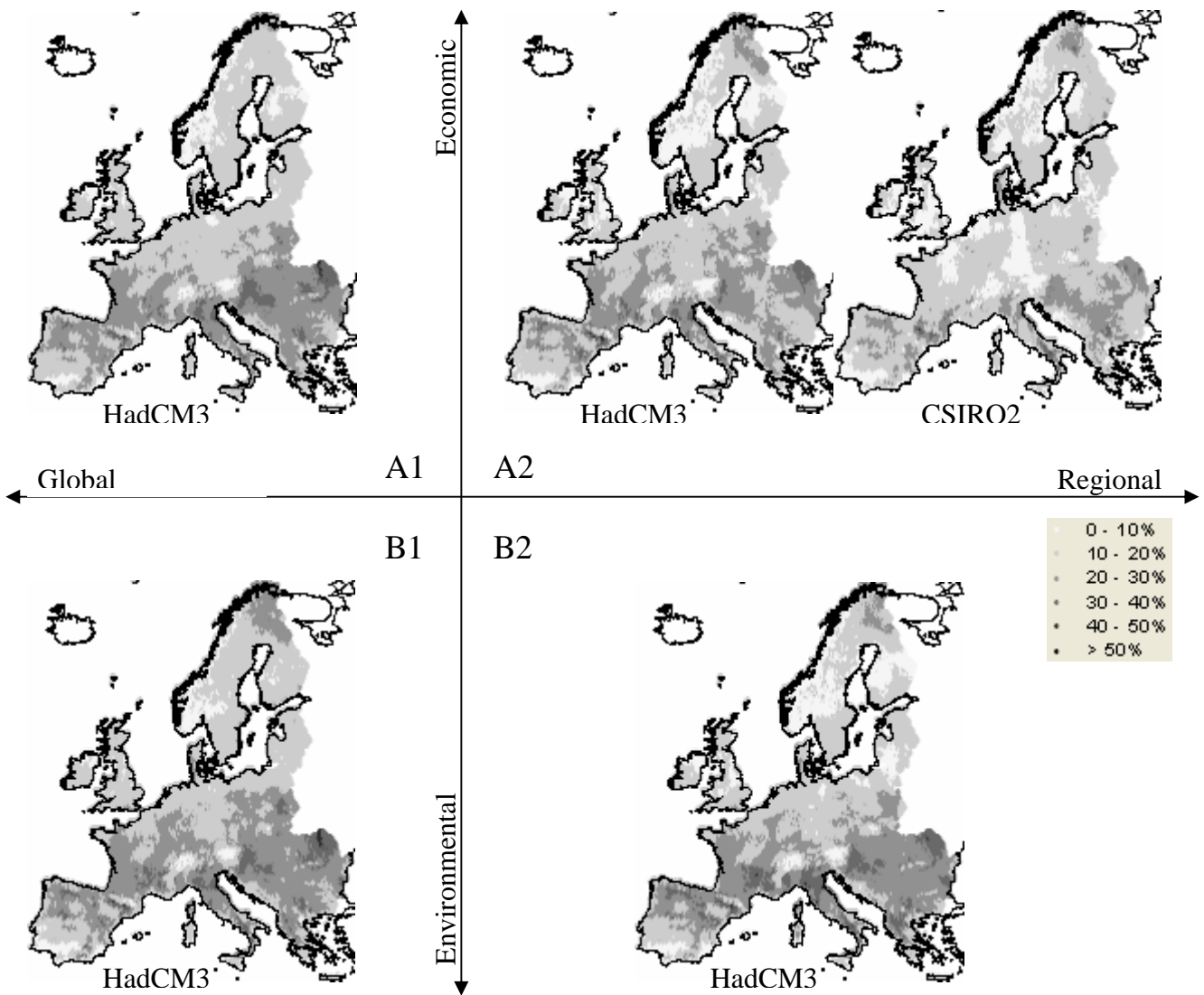
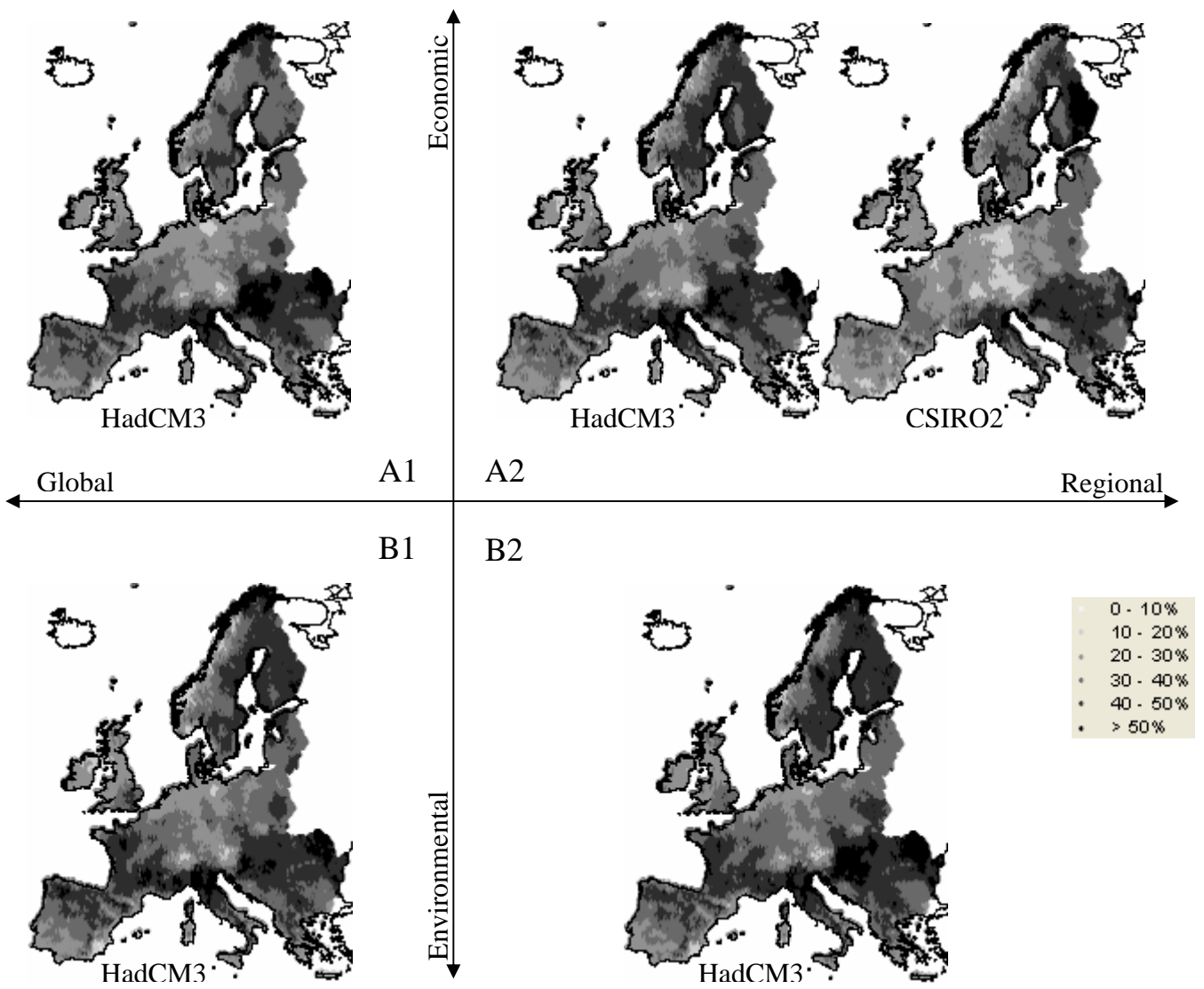


Fig. 3



ANNEXE 3.4

Impact of global climate change on distributions of European taxa: projections under a set of IPCC storylines

Thuiller W., Araújo M.B., Lavorel S., Williams P.H.

En préparation

Impact of global climate change on distributions of European taxa: projections under a set of IPCC storylines

Wilfried Thuiller^{1,2}, Miguel B. Araújo^{1,2}, Sandra Lavorel^{1,3}, Paul Williams⁴

¹Centre d'Ecologie Fonctionnelle et Evolutive. Centre National de Recherche Scientifique. 1919 route de Mende, 34293 Montpellier cedex 5, France.

²Macroecology and Conservation Unit, University of Évora, Estrada dos Leões, 7000-730 Évora, Portugal.

³Laboratoire d'Ecologie Alpine. CNRS. Université J. Fourier. BP 53X, 38041 Grenoble Cedex 9, France.

⁴Biogeography and Conservation Laboratory, The Natural History Museum, London, UK

Because elements of biodiversity respond to climate as an important driving force of distributional limitation^{1,2}, distributional shifts, species loss are expected^{3,4}. Hence, projecting future biodiversity distribution is required to assess the potential impacts of global change and to plan measures to mitigate them.

Nevertheless, despite of modelling efforts on number of species and communities^{5,6} and evidence of climate change-mediated shifts and phenological changes⁷, no have attempted to project single species or assemblages of species alterations according to a range of climate change scenarios supposed to represent the human societal response.

Here we project the future distribution of a sample of European flora and fauna (2,013 species) for 2050, and the potential percentage of expected extinctions and colonisations, according to a range of four storylines on emissions scenarios⁸. Each storyline represent a model society with different demographic, socio-economic and technical responses to change. Although differences by species across the four storylines are not strong, differences for percentage of extinction/colonisation exhibit contrasting patterns in Europe according to the considered country, highlighting which societal response would the most appropriate for preserving biodiversity.

Global climate change is already affecting a broad range of organisms with various geographical distributions^{3,4,7}. To forecast the potential effects of climate change on ecosystems and society, the Intergovernmental Panel on Climate Change (IPCC) developed long-term emissions scenarios. The last set of scenarios based on the Special Report on Emission Scenarios⁸ (SRES) encompasses four combinations of demographic changes, social and economic developments corresponding to the four families A1, A2, B1 and B2 (see Method sections). These scenarios should be considered equally sound and span a wide range of uncertainty and complex interactions between economy, social behaviour and population growth. As these scenarios cover virtually all possible directions of change, from high shares of fossil fuels, oil and gas or coal, to high shares non-fossil, they should be used jointly to project impact of climate change on biodiversity. Several studies have attempted to predict future species distribution at large spatial⁵ or more local scales⁹, but few have analysed the possible impacts of using a panel of scenarios¹⁰. Indeed, as the scenarios are covering a range of human responses, all species are not supposed to respond identically and so according to different scenario, the impact on species distribution should be contrasted.

Using the four storylines of future changes with the GCM HadleyCM3 and new tool in biodiversity

modelling⁽¹⁾ we develop predictions of the effects of global climate change on distributions of 2,013 species (383 birds, 162 mammals, 118 herptiles, 1305 plants)^(10,11). We use the BIOMOD framework, allowing to use four different modelling techniques and select the most consensual one (see Methods section). This first large-scale application of niche-based models, on such number of species to the challenge of predicting the potential consequences of global climate change according to different societal responses, highlight the way to follow to mitigate the impact of global climate change on biodiversity.

To examine patterns of change in distributional area across all 2,013 species, we distinguished local extinction of species by pixel (opposed to remaining species by pixel) and new potential climatically suitable habitat for species (colonisation if species could reach it).

Extinctions were predicted to be heterogeneous across the four storylines (Fig. 1). Storylines A1 and B1 are the most critical to affect species survival with the highest proportion of species loss by pixel. Inversely, storylines A2 and B2, which are supposed to represent a more environmental oriented world, are less critical than A1 and B1, even if some area could reach forty percent of extinction. Predictions from different storylines differed regionally and not in a continental

way making difficult to extrapolate a general pattern across biogeographical regions and country. However, there are areas where model under all storylines predicted similar pattern. This is the case for Alpine mountains which were predicted to undergo little extinction rate (15%), or for southern Spain (12%). Inversely, under all scenarios, northern Spain (Catalonia and south of Pyrenees mountains (dire exactement la région concernée?? Aragon?) and northern of Romania were predicted to loss high rate of species (about 45%).

Similarly to extinctions, predicted colonisations are spatially heterogeneous across storylines. Under A1 storylines, France was predicted to gain between 40-60% of species by pixel, while under the three other scenarios this rate is range from 10 to 40%. Excepted A1, the other storylines exhibit roughly the same pattern with for instance the highest rate of colonisation in Boreal (above 37%) and Pannonian (40%) regions. A part of French Alps mountains were also predicted by all storylines to undergo high rate of expected colonisation (60-70% by pixel). Similarly, Germany was predicted to gain few species by pixel according to A2, B1 and B2 storylines (10-20%) while under A1 model predicted heterogeneous pattern from 10 to 40% of colonisation by pixel).

Nevertheless, percentages of colonisation by pixel are theoretical and represent an unlikely hypothesis of universal dispersal. It is more straightforward to apprehend colonisation as a null hypothesis and the most dramatic situation, being supposed that all species are not able to disperse far and fast enough.

To pursue the analysis of European biodiversity under four storylines, we analyse patterns of extinctions/colonisations by country (Table 1). There was a very high heterogeneity across storylines and across countries. Even if A1 was the most dramatic with the highest mean extinction across Europe, there were countries where the models under others storylines predicted higher extinctions than A1. This is the case for instance for Serbia where under A1 storyline, model predicted a loss of 22% while 22% under B2 storyline. Independently of storylines, all countries in Europe were not susceptible to undergo the same rate of extinctions and colonisations. Indeed, while Hungary was expected to loss 28% of species by pixel in 2050, Norway was expected to loss 11%. Similarly for colonisation, while Hungary was predicted to potentially gain 46% of new species by cell, Germany was only predicted to potentially gain 17%. Finally, extreme northern European country (i.e. Finland, Latvia, Lithuania, Netherlands, Norway and Sweden) and extremes south-eastern and south-western countries (i.e. Greece and Portugal) were the least prone to loss high rate of species by pixel. Inversely, south continental countries (Croatia, France, Hungary, Macedonia, Romania) were predicted to be the most threatened by species extinction.

The analyses presented herein illustrate the magnitude of effects that may be expected as a consequence of

global climate change. Moreover, using different scenarios or storylines delimitate these effects in a range of possible. Refined scenarios as such produce recently for the IPCC allow to test different effects of changes according to different socio-economic responses. High emission of CO₂ coupled with globalisation and economic oriented future world should undergo high rate of extinction and so alter species assemblages, communities and ecosystems functioning. According to our results, B2 storyline encompassing regionalisation and environmental politic and economy is logically the scenario that should not perturb too much species distributions and risk of extinctions.

Of course, projections as carried out in this study can only be used to give insight of the possible effects of climate change on biodiversity because the complexity of nature is highly unpredictable and because models can only summarize a part of exerted environmental forces on species, communities and ecosystems. Nevertheless, supplementary work is required to include dispersal ability, landscape heterogeneity and biotic interactions to refine models predictions to address questions of productivity in economically important ecosystems, emerging diseases and diversity conservation¹².

1. Johnston, K. M. & Schmitz, O. J. Wildlife and climate change: assessing the sensitivity of selected species to simulated doubling of atmospheric CO₂. *Global Change Biology* **3**, 531-544 (1997).
2. Woodward, F. I. *Climate and plant distribution* (Cambridge University Press, Cambridge, 1987).
3. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37-42 (2003).
4. Walther, G.-R. et al. Ecological responses to recent climate change. *Nature* **416**, 389-395 (2002).
5. Bakkenes, M., Alkemade, R. M., Ihle, F., Leemans, R. & Latour, J. B. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology* **8**, 390-407 (2002).
6. Peterson, A. T. et al. Biodiversity consequences of climate change in Mexico. (2002).
7. Root, T. L. et al. Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57-60 (2003).
8. Nakicenovic, N. & Swart, R. (eds.) *Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, 2000).
9. Gottfried, M., Pauli, H., Reiter, K. & Grabherr, G. A fine-scaled model for changes in species distribution patterns of high mountain plants induced by climate warming. *Diversity and Distributions* **5**, 241-251 (1995).

10. Peterson, A. T. et al. Future projections for Mexican faunas under global climate change scenarios. *Nature* **416**, 626-629 (2002).
11. Thuiller, W. BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* **9**, 000-000 (2003).
12. Watson, R. Common themes for ecologists in global issues. *Journal of Applied Ecology* **36**, 1-10 (1999).
13. Jalas, J. & Suominen, J. *Atlas Florae Europaeae* (The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki, 1972-1996).
14. Mitchell-Jones, A. J. et al. *Atlas of European Mammals* (Academic Press, London, 1999).
15. Gasc, J.-P. et al. *Atlas of amphibians and reptiles in Europe* (Societas Europaea Herpetologica & Museum National d'Histoire Naturelle, Paris, 1997).
16. Hagemeyer, W. J. M. & Blair, M. J. *The EBCC atlas of European breeding birds, their distribution and abundance* (Poyser, London, 1997).
17. Araújo, M. B. et al. Would environmental diversity be a good surrogate for species diversity? *Ecography* **24**, 103-110 (2001).
18. Humphries, C. et al. Plant diversity in Europe: *Atlas Florae Europaeae* and WORLDMAP. *Acta Bot. Fennica* **162**, 11-21 (1999).
19. Lahti, T. & Lampinen, R. From dot maps to bitmaps - Atlas Florae Europaeae goes digital. *Acta Botanica Fennica* **162**, 5-9 (1999).
20. Williams, P. et al. Endemism and important areas for representing European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belgian Journal of entomology* **2**, 21-46 (2000).
21. Mitchell, T. D. A comprehensive set of climate scenarios for Europe. *In prep* (2002).
22. Carson, D. J. Climate Modelling: achievements and prospects. *Q.J.R. Meteorol. Soc.* **125**, 1:27 (1999).
23. IPCC. *Climate change 1995: the science of climate change. Contribution of working group I to the second assessment report of the IPCC* (Cambridge University Press, New York, 1996).
24. Hill, J. K., Thomas, C. D. & Huntley, B. Climate and habitat availability determine 20th century changes in a butterfly's range margin. *Proc. R. Soc. Lon. B* **266**, 1197-1206 (1999).
25. Frescino, T. S., Edwards Jr., T. C. & Moisen, G. G. Modeling spatially explicit forest structural attributes using Generalized Additive Models. *Journal of Vegetation Science* **12**, 15-26 (2001).
26. Leathwick, J. R. Climatic relationships of some New Zealand forest tree species. *Journal of Vegetation Science* **6**, 237-248 (1995).
27. Rouget, M. et al. Determinants of distribution of six *Pinus* species in Catalonia, Spain. *Journal of Vegetation Science* **12**, 491-502 (2001).
28. Thuiller, W. et al. Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology & Biogeography* **12**, 313-325 (2003).
29. Peterson, A. T. Projected climate change effects on Rocky Mountain and Great Plain birds: generalities of biodiversity consequences. *Global Change Biology* **9**, 647-655 (2003).

Figure 1 Modelled species extinction rate by pixel for 2013 species across Europe. A) under A1 scenario; B) under A2 scenario; C) under B1 scenario; and D) under B2 scenario

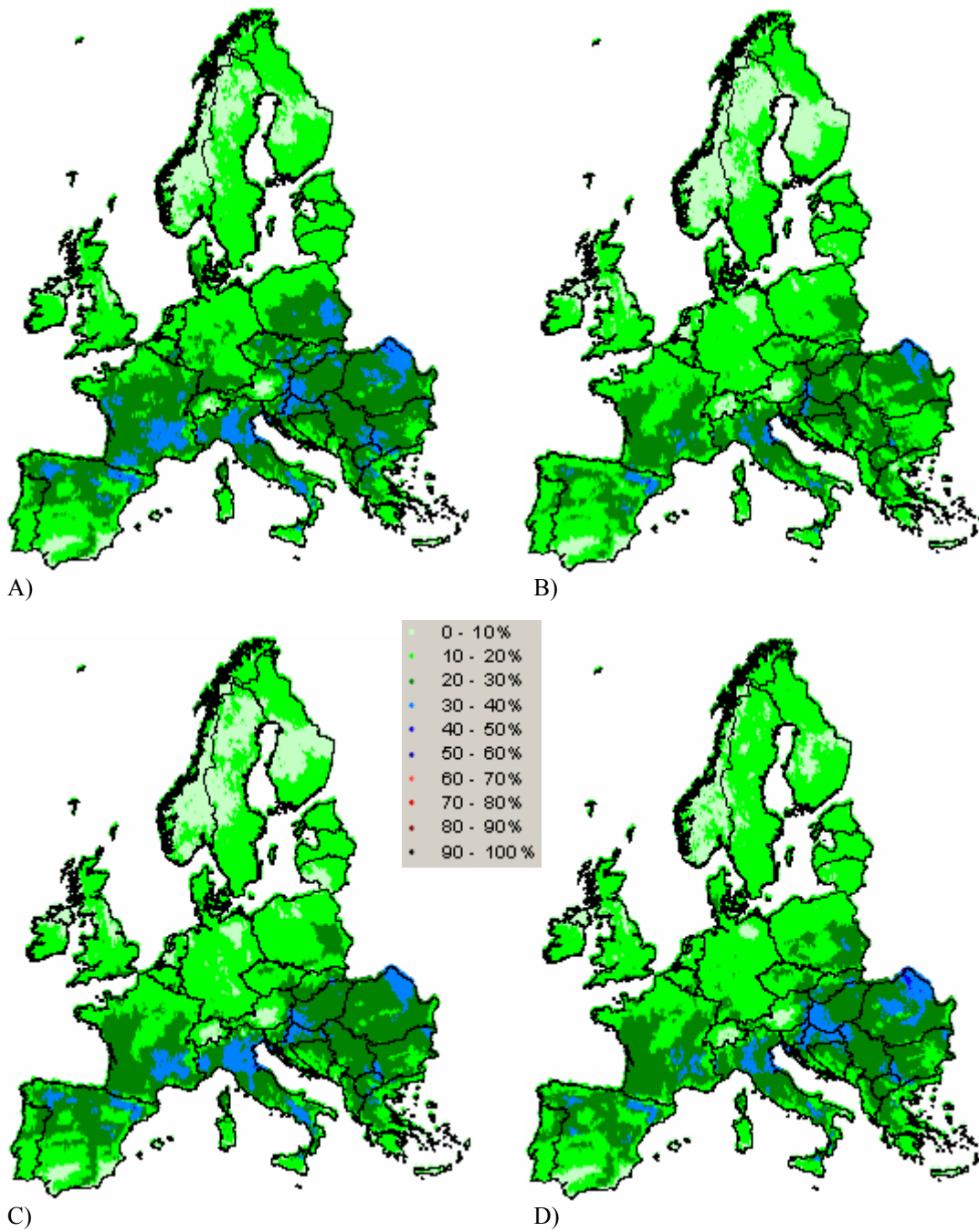


Figure 2 Modelled species colonisation rate by pixel for 2,013 species across Europe. A) under A1 scenario; B) under A2 scenario; C) under B1 scenario; and D) under B2 scenario.

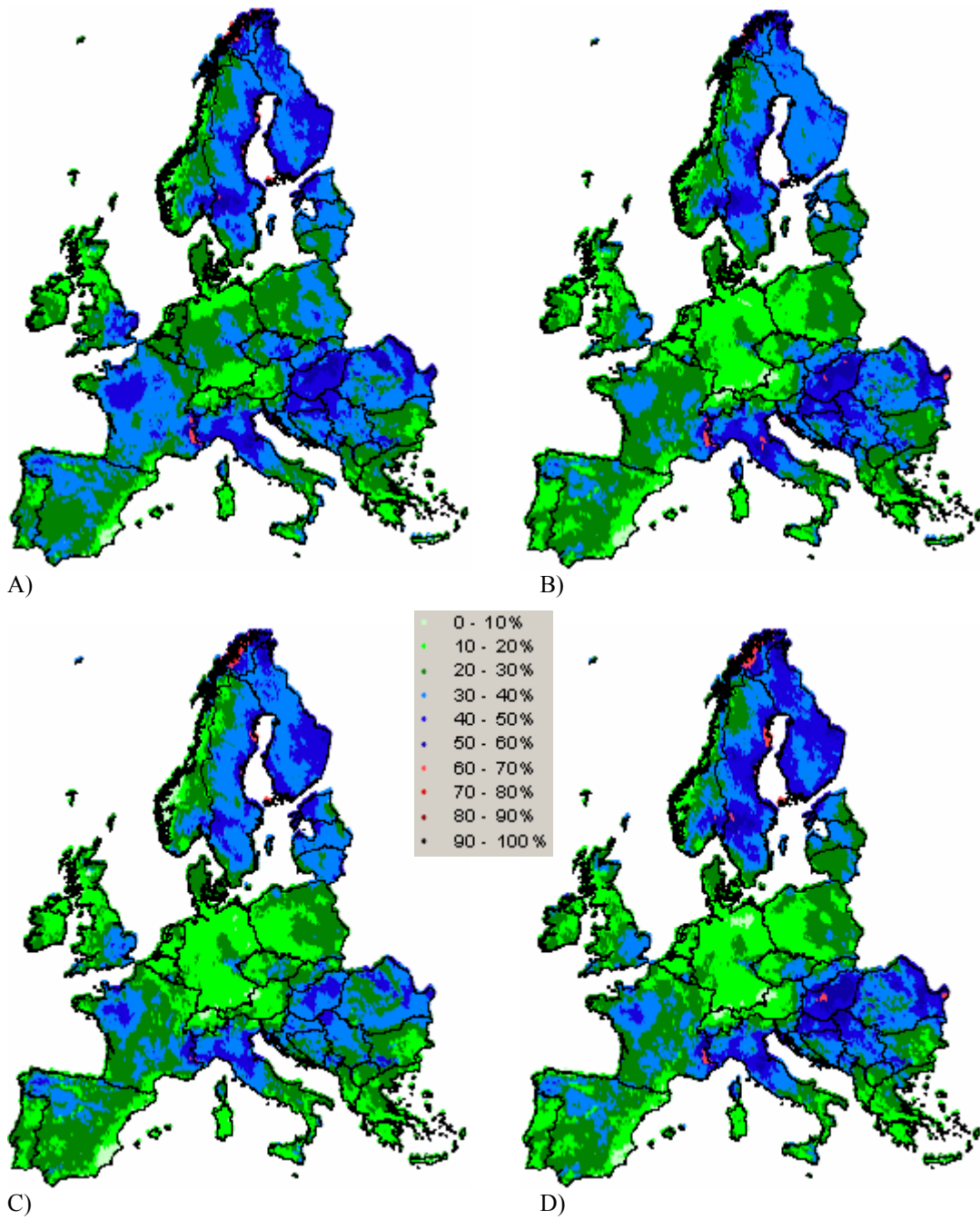


Table 1: Mean expected percentage of species' extirpations and colonisations in European countries according to four storylines of climate change. Ext and Col represent respectively the expected mean percentage of species lost by cell and mean percentage of species gained by cell. Bold numbers represent the minimum and maximum values of extinction and colonisation by storylines

Country	A1		A2		B1		B2	
	Ext	Col	Ext	Col	Ext	Col	Ext	Col
Albania	19.8	25.6	15.4	27.1	20.8	21.7	18.9	24.2
Austria	19.2	22.4	14.8	19.5	16.0	18.1	17.2	19.1
Belgium	17.9	24.2	14.4	19.7	14.6	18.7	15.9	19.2
Bosnia H.	23.3	36.7	20.4	39.2	22.2	31.5	23.3	35.9
Bulgaria	24.1	27.8	19.6	29.1	23.7	25.6	23.2	28.9
Croatia	26.9	40.1	23.7	42.6	26.3	33.1	26.7	40.7
Czech R.	22.9	32.1	17.8	24.7	17.3	22.9	18.6	24.4
Denmark	18.2	23.0	15.7	21.3	15.3	23.3	16.4	21.2
Estonia	17.5	38.2	15.6	32.9	15.2	38.6	15.6	35.3
Finland	13.0	40.6	11.7	36.8	11.7	39.6	13.4	42.5
France	24.2	33.8	20.0	27.2	21.7	28.0	21.4	28.4
Germany	17.7	23.1	13.5	16.5	13.3	16.3	14.9	15.7
Greece	17.2	21.8	14.5	21.5	17.1	19.0	17.2	20.8
Hungary	28.0	46.5	23.7	47.7	27.2	37.2	28.1	49
Ireland	12.9	20.5	11.3	19.7	12.2	19.6	12.7	21.9
Italy	23.1	33.3	20.6	34.1	23.8	30.0	22.0	32.3
Latvia	16.0	33.7	14.2	29.8	13.9	34.8	14.4	30.2
Lithuania	13.4	29.9	11.3	26.8	10.8	32.6	12.0	26.6
Luxembourg	22.8	27.3	18.2	22.8	18.0	21.4	19.3	22.7
Macedonia	27.1	28.0	21.3	28.5	25.5	25.2	25.0	28.3
Montenegro	18.4	37.6	15.1	42.2	17.8	33.1	18.6	36.9
Netherlands	12.6	21.8	10.7	18.8	10.9	19.1	12.2	18.4
Norway	11.0	29.9	10.8	29.1	11.7	29.1	12.7	33.0
Poland	20.6	27.2	16.5	21.4	15.5	21.1	17.9	20.9
Portugal	17.6	24.6	16.3	19.8	18.0	21.2	17.3	20.1
Romania	26.8	37.8	22.9	37.5	25.8	32.6	27.5	38.5
Serbia	22.1	39.8	25.2	32.8	26.3	39.9	25.7	35.4
Slovakia	25.7	35.4	21.5	32.4	21.5	27.0	23.6	32.2
Slovenia	23.5	28.2	20.7	30.6	20.9	22.1	23.1	28.1
Spain	19.9	26.9	18.4	23.0	20.4	24.1	19.7	24.0
Sweden	12.1	35.4	11.0	33.5	11.2	32.7	12.5	37.8
Switzerland	16.6	21.5	13.1	21.6	13.0	19.0	13.9	19.7
UK	13.5	23.5	12.2	21.5	12.9	21.7	13.1	22.9
<i>Mean</i>	19.6	30.2	16.7	28.2	18.0	26.7	18.6	28.3
<i>St dev</i>	4.9	6.9	4.3	7.9	5.2	7.0	4.9	8.2

METHODS

Data on current species distributions:

Original species data include 2,294 plant⁽¹³⁾, 186 mammal⁽¹⁴⁾, 143 amphibian and reptile⁽¹⁵⁾ and 440 breeding bird species⁽¹⁶⁾. Terrestrial vertebrates include all known species⁽¹⁷⁾ whereas plants comprise only ~20% of the European flora⁽¹⁸⁾. The grid used was based on the *Atlas Florae Europaeae*⁽¹⁹⁾, with cells boundaries typically following the 50 km lines of the Universal Transverse Mercator (UTM) grid. The mapped area comprises western, northern and southern Europe, but excludes most of the eastern European countries where recording effort was both less uniform and less intensive⁽²⁰⁾. After removing species with less than twenty records necessary for model accuracy, the data set comprised 1305 plants, 162 mammals, 118 herptiles and 383 birds.

Current and future climate data:

Climate data used for analyses were a comprehensive set of bioclimatic variables for Europe from the Climatic Research Unit (<http://www.cru.uea.ac.uk/>)²¹ (<http://www.pik-potsdam.de/ateam/>): mean annual, winter and summer precipitation, mean annual temperature and minimum temperature of the coldest month, growing degree days (>5°) and an index of humidity (mean ratio of annual actual over annual potential evapotranspiration). Means were averaged for the period of 1961-1990. These data sets were supplied on a 10' grid, covering Europe. Then data were aggregated by averaging to 50x50 km UTM in order to match the resolution of species data. Niche-based models were calibrated on the 50x50 km UTM grid and then species distributions projected on the 10' grid for current and future climate.

Future projections were derived using two different general circulation model for period 2050²²: The GCM experiments (HadCM3) make available through the IPCC-DDC (Data Distribution Centre of the IPCC: ipcc-ddc.cru.uea.ac.uk) and they are conducted at the UK Hadley Centre for Climate Prediction and Research (HadCM3). These GCM included four storylines scenarios according to IPCC (Intergovernmental Panel on Climate Change)' recommendations⁸: The A1 storyline describes a future world of very rapid economic growth, global population that peaks in mid-century and decline after, and the rapid introduction of new and more efficient technologies. Concentration of CO₂ should pass from 380 ppm in 2000 to 580 ppm in 2050 and temperature should increase of 2°C⁽²³⁾. The A2 storyline describes a very heterogeneous world. The underlying theme is self-reliance and preservation of local identities. Economic development is primarily regionally oriented and per capita economic growth and technological change are more fragmented and slower than the other storylines. Concentration of CO₂ should pass from 380 ppm in 2000 to 520 ppm in 2050 and temperature should increase of 1.5°C⁽²³⁾. The B1 storyline describes a convergent world with the same

global population that peaks in mid-century and declines thereafter, as the A1 storyline, but with rapid change in economic structures toward a service and information economy, with the introduction of clean and resource-efficient technology. Concentration of CO₂ should pass from 380 ppm in 2000 to 490 ppm in 2050 and temperature should increase of 1.2°C⁽²³⁾. The B2 storyline describes a world in which the emphasis is on local solutions to economic, social, and environment sustainability. It is a world with continuously increasing global population at a rate lower than A2, intermediate levels of economic development, and less rapid and more diverse technological change than in the B1 and A1 storylines⁽⁸⁾. Concentration of CO₂ should pass from 380 ppm in 2000 to 490 ppm in 2050 and temperature should increase of 1.5°C⁽²³⁾.

Niche-based models of species climatic envelopes:

Niche-based models relate the present-day distribution of species to a pool of available environmental variables using statistical or rule-based methods. Different modelling techniques are commonly applied to project potential future species distributions. Although Generalised Linear Models (GLM) are the most common^(5,24), Generalised Additive Models (GAM) are being used increasingly^(25,26). Alternative rule-based approaches have also emerged as interesting tools to predict current and future potential distributions. These include Classification And Regression Tree analysis (CART)^(27,28), Artificial Neural Networks (ANN) (Pearson, 2002 #1379) or Genetic Algorithms^(10,29).

We used BIOMOD, a framework which capitalizes on the different techniques (GLM, GAM, ANN, CTA) used for niche-based modelling to provide alternative spatial predictions⁽⁶⁾. For each scenario, we run the four niche-based models and select the most consensus one. The most consensus projection was selected as the most related to the first axis of a factor analysis carried out on the four species-projections. This allows to select the model' projection the most agree with the others and so the more conservative one.

Percentage of extinction/colonization/species turnover:

Each species was predicted to loose or gain suitable habitat in the future. To evaluate the percentage of extinction, we summed the number of species lost (L) by pixel (opposed to stable species) and related it to current species richness by pixel. The procedure was the same to assess the percentage of species gained (G) by pixel (under assumptions that species could reach new habitat).

ANNEXE 4.1

Consequences of land use change on rare plant occurrence and extinction in a Mediterranean region: a 115 years study

Lavergne S., Thuiller W., Molina J., Debussche M

En préparation

Consequences of land use change on rare plant occurrence and extinction in a Mediterranean region: a 115 years study

Sébastien Lavergne^{1*}, Wilfried Thuiller¹, James Molina² and Max Debussche¹

¹Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, 1919 route de Mende, F-34293 Montpellier cedex 5, France

²Conservatoire Botanique National Méditerranéen de Porquerolles, Antenne Languedoc-Roussillon, 163 rue Auguste Broussonnet, 34090 Montpellier, France

* *Corresponding author* :

Sébastien LAVERGNE,

University of Vermont, Department of Botany and Agricultural Biochemistry,

Marsh Life Sciences Building, 109 Carrigan Drive, Burlington, Vermont 05401-0086 USA

sebastien.lavergne@uvm.edu

Tel : 802-656-0430 / Fax : 802-656-0440

ABSTRACT

During the 20th century, Mediterranean regions have been subject intense and complex land use changes, and are thus supposed to be exposed to important loss of biodiversity. In this study, we studied rare plant species distribution in a 6,250 sq km region of the French Mediterranean region, using two botanical surveys conducted in 1886 and in 2001. We developed empirical models to explain spatial patterns in rare species richness in 1886 and 2001 with environmental and human variables. The 1886 distribution of rare species occurrences is mainly explained by the spatial variation of the 1886 prospecting intensity. Most of regionally rare species are present in high altitudes in zone of extensive agriculture, dominated by semi-natural open habitats, and where agriculture has been regressing between 1886 and 2001. However, the different species rare in the region in 2001 can even exhibit contrasted responses to ecological and human factors, because of varied biogeographical origins and because of different

histories of population dynamics in the region. Eurosiberian and Mountain rare species are the most marginal in the region and seem to incur a high risk of regression in the Mediterranean under the current predictions of global change. The population of rare species are more likely to persist if they occur in zones where harsh topography and low fertility may limit accessibility and exploitation: many restricted endemic rare species seem to be in that case. On the contrary, regionally rare species are the most prone to extinction when occurring in zones of current urbanization and agriculture intensification. We hope these results to provide a solid scientific basis for a redistribution of conservation efforts among rare species of the Mediterranean regions.

KEYWORDS: rarity, extinction, regional scale, Mediterranean region, environmental factors, land use change, chorology, endemism, outlying mean index, generalized linear models.

INTRODUCTION

The worldwide human impacts on earth ecosystems are currently supposed to cause dramatic losses of species from local to global scale due to habitat modifications, destructions, fragmentations, biological invasions and climate change (Parmesan & Yohe, 2003; Pimm, 1996; Vitousek et al., 1997). Identifying the geographical patterns of rare or endangered species is thus crucial for understanding how human activities can threaten biodiversity, and ultimately for prioritising sites or species to preserve (Prendergast et al., 1993a). Hotspots of biodiversity and rarity have been accurately matched at global or continental scale (Dobson et al., 2001; Myers et al., 2000). However, in order to understand the mechanisms by which human activities and land use

changes can affect rare species persistence or extinction, more spatial studies must be achieved at mesoscale, i.e. on small biogeographic region or provinces (e.g. Heikkinen 1998; Luoto 2000).

The Mediterranean Basin is one these regions where future land use and climatic scenarios reveal important threats to large amounts of biodiversity (Sala et al., 2000). The Mediterranean flora consists of 25,000 vascular plant species, whose about 59% are considered as narrow endemic species (Greuter 1991). Regional assemblages of plant species result from complex interactions between geological history and climatic influences: many Eurosiberian taxa reach their range limits in the Mediterranean regions and co-occur with Southern-Central Europe taxa, Mediterranean taxa, or Mountain taxa (Quézel, 1985). Since 8,500 yr

B.P., Mediterranean ecosystems had been continuously shaped by human activities such as extensive agriculture and animal breeding (Pons & Quézel, 1985), but the intense land use changes that occurred during the 20th century have had profound consequences on most Mediterranean landscapes (Lepart & Debussche, 1992). Backcountries have been subject to a dramatic rural depopulation and to a decrease in traditional extensive farming activities, resulting in the spread of woodlands by secondary succession, and in the loss of open habitats such as semi-natural grasslands and heathlands (Debussche et al., 1999). On the contrary, rapid urbanisation occurred on seashores and at the periphery of main cities and, in lowlands, agriculture has been substantially intensified following the second World War (Pons & Quézel, 1985).

A critical task in the Mediterranean region is thus to identify areas of land where rare species are at greatest risk of extinction and to determine the underlying mechanisms of these threats. Nevertheless, the opportunities to empirically assess the long term impacts of land use changes on rare species are very scarce because comparative data sets on species presence or abundance are required over long periods of time. The use of archival biological information could thus help to improve our knowledge on human-induced extinction processes (e.g. Fischer & Stöcklin 1997; Duncan & Young 2000). Moreover, at regional scales, plant species can differ in their response to environmental factors according to their chorological status (Thuiller et al., 2003) and some authors stressed that pools of rare species could not be considered as an ecologically consistent assemblages because of different histories of rarity (Fiedler, 1986). Regionally rare plant species can thus exhibit various response groups to environmental and human factors (Hill & Keddy, 1992), because of varied biogeographical origins and different histories of population dynamics in the region.

In this study, we used a 115-year historical recording of rare plant species distribution in a Western Mediterranean region to assess the effect of environmental factors, human activities and land use changes on the distribution and fate of rare plant species over the 20th century. Here, rarity is measured as the number of occurrences recorded in 1886 and in 2001 within the “Hérault Département”, a 6,250 km² French administrative circumscription. In 1886, Loret & Barrandon surveyed the whole flora of the “Hérault Département” and provided locations of the rarest species; i.e. present in or two sites in the study area. In 2001, the French National Mediterranean Botanical Conservatory achieved the same survey, which also allowed to check the current status of rare species occurrences in 1886 (extinct, stable) and to determine the species appeared on the “Hérault Département” from 1886 to 2001. We then developed empirical models to relate spatial patterns of rare species richness in 1886 and 2001, and of rare species dynamics between 1886 and 2001 (extinct, stable,

appeared) with environmental variables (geological and topographical), human population and land use variables. We addressed the three following questions:

- (i) Do environmental, human and land use factors affect spatial pattern of rare and/or endemic species richness in 1886 and 2001?
- (ii) Are these effects different between chorological classes of rare species, i.e. Mediterranean, Southern / Central Europe, Mountain and Eurosiberian species?
- (iii) Do environmental factors and land use changes affect fate of rare species occurrences in our region (extinction, stability, appearance) from 1886 to 2001?

MATERIAL & METHODS

Study area

The Hérault Département is a French administrative division extending on 6,250 km², from the Mediterranean seashore to the southern mountain ranges of Massif Central (Fig 1). Its vascular flora is composed of ca. 2 400 native species of phanerogams and ferns, i.e. 55 % of the whole flora of France on only 1.1 % of the French territory. The climate is mainly Mediterranean, with a dry and warm summer and precipitations mostly occurring, sometimes in huge amounts, in autumn and in spring. However, the north-western and highest part of the Hérault Département experience some Oceanic influences (Daget, 1977). Frost occurs each year, even near the coast. Important gradients of rainfall (400 – 1,600 mm/yr) and temperature are observed with distance from the sea and towards higher elevation (i.e. up to 1205 m a.s.l.). Bedrocks are of diverse origins. Sedimentary bedrocks (limestone, dolomite and marl) mainly occur at low altitude and on the north-eastern part of the Département, whereas metamorphic (schist) and intrusive (granite) bedrocks mainly occur on the north-western and highest part (Dugrand 1971). Though localised, the presence of basalt outcrops must be underlined, as they harbour particular habitats (e.g. peatbogs, temporary ponds) of special importance for several rare plants.

As well as in the rest of the Mediterranean region, landscape of the Hérault Département reflects a long and complex interactions between human activities and environmental factors. Four main zones can be distinguished along a broad south-north transect, in relation with their altitude, land-use, urbanization and vegetation type dominance: the littoral, the plain, the hills and piedmont, the plateaux and mountains (Le Floc'h et al., 1983).

- In the littoral zone, tourism facilities and urbanized area have also been overwhelmingly increasing since the 70s, along the seashore and at the periphery of small cities. Vineyards and orchards have been extended on large areas by draining most marshes and wet grasslands. Natural and semi-natural habitats consist of dunes, lagoons, marshes, salt or humid meadows, with scattered evergreen shrublands and woodlands of Aleppo Pine (*Pinus halepensis*).

- The plain zone has been subject to a rapid urbanization and to intensification of agriculture

(mainly wine growing) since the 60s. Evergreen shrublands (*Quercus coccifera*, *Cistus monspeliensis*, etc.) and woodlands of Aleppo Pine and evergreen Holm Oak (*Quercus ilex*) are localized on rocky outcrops.

- The hills and piedmont have severely suffered from rural depopulation. Agriculture has almost vanished and flocks of sheep, counted by hundreds of thousands of animals one century ago, have disappeared. Woodlands of Holm Oak or deciduous Downy Oak (*Quercus pubescens*) and old fields colonized by shrubs (*Genista scorpius*, *Juniperus oxycedrus*, etc.), are the main components of the landscape, and illustrate how plant succession following this abandonment has developed rapidly. On acid soils and upper sites, Chestnut trees (*Castanea sativa*) and heathlands (*Erica scoparia*, *Calluna vulgaris*, etc.) are frequently observed.

- The plateaus and mountains have also been subject to a strong rural depopulation, but breeding and agriculture have persisted from place to place. Limestone plateaus are covered by dry grasslands, being currently invaded by shrubs (*Buxus sempervirens*, *Juniperus communis*, etc.), and low forests of deciduous Downy Oak and Sessile Oak (*Quercus sessiliflora*). The Mountains, granitic and schistous, are mostly covered by native woodlands of Sessile Oak, Chestnut and Beech (*Fagus sylvatica*), and non-native coniferous forests (*Pinus laricio*, *Picea excelsa*) with patches of grasslands and heathlands (*Cytisus purgans*, *Sarothamnus scoparius*, *Pteridium aquilinum*, etc.).

Species data sets

The Lorret & Barandon (L&B hereafter) survey of 1886, and the French National Mediterranean Botanical Conservatory survey of 2001 respectively recorded the presence of 2,038 and 2,162 plant species in the Hérault Département. This study focused on species recorded as rare in the study region (i.e. one or two sites) in 1886 and/or in 2001. This threshold of two occurrences in the study region was not chosen arbitrarily. As the occurrences locations of the rarest species in 1886 were precisely reported only for species present in one or two sites, the same criteria was applied in the 2001 survey to delimitate the actual sample of rare species. Moreover, this criteria fits to the ones commonly used to elaborate lists of regionally protected species (Molina et al., 1999).

The study also included only taxa of species rank at both dates according to Flora Europae (Tutin et al., 1964-1993). Since taxonomic nomenclature has changed from 1886 to 2001, we excluded species for which nomenclature confusion biased the abundance changes estimation between 1886 and 2001. To do so, we translated the nomenclature used by L&B and check the actual species rank of the taxa described by L&B in 1886. We used the French Flora Synonyms Index (Kerguélen, 1993), and the original herbarium samples of L&B (stocked in Botanical Institute of Montpellier) when necessary.

The final lists of rare species sites concerned 417 species: 143 rare species for 1886 and 219 rare species for 2001 (i.e. 9.9% of the overall flora of Hérault in 2001). Rare species occurrences in 1886 and in 2001 were georeferenced under ARCVIEW GIS 3.2a (ESRI Inc., Redlands, California), using digitised 1/25,000 maps generated by French National Geographic Institute. The comparison of the 1886 and 2001 list of rare species and of their occurrences allowed us to determine which occurrences of 1886 rare species remained stable or were gone extinct between the two dates, and which occurrences of 2001 rare species have appeared since 1886.

To take into account the distribution of species outside the study area, we determined for each species its geographical extent and its chorological affinity (equivalent to phytogeographic groups), following Coste (1900-1906), Tutin *et al.* (1964-1993), and Greuter *et al.* (1984-1989). Geographical extent was coded as restricted endemic or widespread species. Chorological status of each species was determined from their geographical distribution with the following classes:

- Mediterranean and Sub-Mediterranean species: species distributed within the Mediterranean region and hardly extending beyond its borders. For some of these species, the Hérault Département is the absolute northern limit of their distribution.
- Southern / Central Europe species: species mostly distributed throughout southern and central Europe, and /or parts of North Africa and Western Asia, and absent from northern Europe.
- Mountain species: species mostly distributed in mountains of southern Europe and possibly in boreal Europe.
- Eurosiberian species: species largely distributed throughout Europe and often present in Asia, but generally rare in the Mediterranean region where they can occur in alpine or humid habitats. The Hérault Département often constitutes a local southern limit of their distribution.

Environmental, human and land use variables

Spatial variation in sampling effort may introduce bias in the study of species distribution (Prendergast et al., 1993b). As it is unlikely that L&B uniformly prospected the study area, Denelle *et al.* (1995) have quantified their prospecting intensity within the 11 main ecological zones recognized in the Hérault Département. This information was georeferenced to provide an explanatory variable of recording effort for 1886 distribution of rare species. Since the 2001 survey of the National Mediterranean Botanical Conservatory has been achieved in contact with several experienced field botanists throughout the study area, we assumed the prospecting effort to be homogeneous in 2001.

Main hypotheses of the study concerned effects of human populations and land-use on spatial patterns of rare plant species richness, but environmental

factors are also likely to influence the occurrence and persistence of rare taxa. We selected bed-rock, elevation and slope (Table 1) because these environmental variables are known to broadly influence plant species distribution at mesoscale (Heikkinen, 1998; White & Miller, 1988), because they are generally highly correlated with many other physical variables, such as temperature or precipitations. We adapted and digitised the 9 following bed-rock categories from Dugrand (1969): Alluvial / colluvial depositions, Limestone, Dolomites / Chert limestone, Marly limestone, Sandstone / Marl, Shales / Quartzite, Gneiss / Mica-schist, Granite, and Basalt. Elevation and slope were derived from a Digital Elevation Model (50m resolution, provided by SIG Languedoc-Roussillon and Cartography Office of Region Languedoc-Roussillon).

To quantify spatial distribution of human populations and changes in this distribution, we digitised spatial data of human population density within the 343 Communes of Hérault (communes size from 1.5 to 12 km²), using the National Demographic Census (INSEE, Montpellier). Commune boundaries were used to calculate human population density in each commune in 1881, 1954 and 1999, and to estimate the changes in human population density from 1881 to 1999 and from 1954 to 1999. This was conducted using ARCVIEW GIS (see Table 1 for calculations). We thus dispose of explanatory variable which quantify both the human impact on given dates but also the changes of this impact over the last 100 years.

As comparable agricultural land-use data was unavailable over the dates considered above, we used data on agricultural practices for 1970 and 2000 provided by the Agricultural National Census (RGA, Agreste, France). For these two dates, we used three synthetic variables to describe agricultural practices in each of the 343 communes of Hérault: (i) permanent grassland and heatland area, (ii) total cultivated area and (iii) total livestock (number of ovine, caprine and bovine heads). Under ARCVIEW GIS, these variables were normalized by the commune area (Table 1) to quantify: (i) the proportion of land devoted to open and extensive semi natural habitats, (ii) the proportion of land modified by agriculture, and, (iii) the density of livestock then the intensity of livestock grazing. The changes in these variables from 1970 to 200 were also calculated to quantify the extent of recent land use changes (Table 1)

Numerical analyses

The study of spatial pattern of rare plant species occurrence and extinctions was performed through a grid of 1x1 km cells. In each grid square, we calculated mean value of each quantitative explicative variable (listed in Table 1) and the dominant lithology classes with ARCVIEW GIS 3.2a (Spatial Analyst 2.0a and personal scripts routines).

For each cell we counted the occurrence of rare species to obtain the four following response variables:

number of occurrences of rare species across Mediterranean, Central/Southern Europe, Mountain, and Eurosiberian chorotype. Then, we performed an Outlying Mean Index (OMI) analysis on the four distinct chorological group (Dolédéc et al., 2000). OMI is a multivariate co-inertia analysis especially dedicated to separate ecological niche across group of occurrences in a multidimensional space defined by the environmental variables. Here, we used the OMI to approximate the “niche” occupied by each chorological group of rare species. We then tested if the four different chorological classes of rare species had different response to environmental, human and land use variables. To do so, the OMI analysis estimate how marginal are the mean ecological conditions used by each chorological group respecting to the mean conditions of the study area (Choler & Michalet, 2002). A normalised PCA analysis was first performed to determine the major environmental gradients. Second, the mean habitat condition used by each chorological group of species was projected on this environmental hyperspace. OMI was performed using the ADE4 library (Thioulouse et al., 1997) implemented into R software (Ihaca & Gentleman, 1996), .

In the 1x1 km grid cells, we also counted the following response variables: (a) the number of rare species occurrences in 1886, (b) the number of rare species occurrences in 2001, (c) the number of rare species local extinction from 1886 to 2001, (d) the number of rare species occurrences stable from 1886 to 2001, (e) the number of appeared rare species occurrences from 1886 to 2001, (f) the number of restricted endemic species occurrences in 1886 and, (g) the number of restricted endemic species occurrences in 2001. The response variables naturally follow Poisson distributions, but their distributions are highly skewed toward 0 (many cells with no rare species occurrences). Then the response variables were modeled with General Linear Models (GLM hereafter) with Negative Binomial error (Venable & Ripley, 2002), using Splus software (Anon, 1999). We fitted two kinds of GLM: GLM with linear terms only (GLMs hereafter), and GLM with linear and polynomial terms (GLMp hereafter). The additional advantage of GLMp is that polynomial terms allow skewed or bimodal responses (Austin & Gaywood 1994). Since potential explicative variables were numerous and potentially inter-related, we used a stepwise selection procedure to determine the environmental, human and land use variable that explain the best spatial patterns of rare species occurrence and extinction. After stepwise selection, GLMs and GLMp were compared by performing a sequential likelihood ratio tests for Negative Binomial generalized linear models (Venables & Ripley, 2002). The best model retained had the highest explained deviance.

RESULTS

The first two axes of the OMI analysis captured 95.6 % of variation in rare species niches. Ordination diagrams on the first two axes of the OMI describes the environmental gradients which best discriminated the occurrences of the four chorological group of rare species (Fig 2). The first axis separates zones of high elevation, with acid substrates (gneiss and mica-schist) from lowland zones with bedrocks of sedimentary origin (alluvial or colluvial depositions, sandstones, marls and marly limestones) and with high percentage of cultivated area and population density (Fig 2a). The second axis of the OMI analysis is only related to the occurrence of basalt bedrocks, which showed no correlation with any other environmental variables (Fig 2a).

These two main ecological gradients do not have the same impact on the spatial distribution of rare species occurrences according to their chorotype. Mediterranean and Southern / Central Europe rare plant species tend to occupy non-marginal habitats in the study area: the centre of the first plane of the OMI analysis is included in the envelope described by their occurrences (Fig 2b, 2c). However, Mountain and Eurosiberian rare species are highly discriminated from the others chorotypes along the first OMI axis. They both tend to occupy elevated zones of the study area, on gneiss and mica-schist substrates (Fig 2d, 2e). However, Eurosiberian rare species also seem to constitute a heterogeneous group, with some of these species niches preferentially occurring on basalt substrate, i.e. along the second OMI axis (Fig 2e).

For the analysis of spatial patterns of rare species occurrences, appearance, stability and extinction, stepwise selection procedures always retained two or more explanatory variables in final GLM (Table 2). Here, we present the four most explicative variables, which usually explained more than 90 (with highest explained deviance). GLMp were always preferred to GLMs except to model the number of rare species extinction (Table 2). To interpret the results, we plotted the response curves of the response variables against the most explanatory environmental, human and land use variables.

Rare species richness in 1886 increased along with mean altitude (Fig 3a): almost no rare species in 1886 occur at under an elevation of 400m. Prospecting intensity of L&B also explained a substantial part of spatial pattern of rare species richness in 1886 (Fig 3b): rare species were more numerous in the most prospected areas. Highest rare species richness in 1886 were reached for intermediate values of 1881 population density (Fig 3c), and on basalt, shale / quartzite, gneiss / mica-schist and limestone substrates (Fig 4a). Rare species richness in 2001 has a slightly different relation with mean altitude (Fig 3d): most rare species occur upper than 700m elevation, and a little part under 200m, but almost no rare species are found at intermediate elevations (Fig 3d). Rare species richness in 2001 also increased with grassland and heat-land area (Fig 3e), but decreased with the

variation of cultivated area from 1970 to 2000 (Fig 3f). This latter result indicates that the fewer rare species are found in area where cultivated area has the more increased in the recent years. Rare species of 2001 were the most numerous on basalt, dolomite / chert limestone, shales / quartzite, limestone, and alluvial / colluvial depositions (Fig 4b).

To explain spatial variation in number of endemic species occurrences in 1886, stepwise selection procedure retained only mean slope and intensity of prospecting (Table 2). Since endemic species richness in 2001 responded similarly to mean slope, we only present the results of analysis of endemic species richness in 2001 (Fig 5). Probability of endemic species occurrences in 2001 increased along with mean slope angle (Fig 5a), but decreased drastically with population density in 1999 (Fig 5b) and with the variation of cultivated area from 1970 to 2000 (Fig 5c).

The spatial variation in richness of rare species occurrences appeared between 1886 and 2001 was best explained by the prospecting intensity of L&B. Most of rare species occurrences appeared from 1886 to 2001 were present in zones of low sampling effort in the 1886 survey. Then, the biological interpretation of appeared occurrences of rare species seems dangerous and we further considered only the spatial patterns of extinction and stability of rare species occurrences from 1886 to 2001. Three land use variables contribute positively to probability of rare species extinction from 1886 to 2001: variation of livestock density from 1970 to 2000 (Fig 6a), variation of cultivated area from 1970 to 2000 (Fig 6b) and variation of population density from 1881 to 1999 (Fig 6c). Rare species extinctions were more numerous on gneiss/mica-schist, basalt, limestone and alluvial/colluvial depositions (Fig 4c). However, rare species were more likely to have stable dynamics between 1886 and 2001 if they occurred in habitats of higher altitude (Fig 7a), of higher slope (Fig 7b), and on basalt, limestone, and marly limestone (Fig 4d). Richness of stable rare species occurrences increased very slightly in zones of low percentage of cultivated area (Fig 7c).

DISCUSSION

A number of methodological difficulties may bias the richness patterns of rare plant species. For example, the compilation of archival data plagues the important problem of prospecting intensity (Prendergast et al., 1993b). Because of limited handling opportunities in the late XIXth century, L&B survey may have been spatially heterogeneous. Hence, the 1886 distribution of rare species and endemic species occurrences is clearly influenced by the spatial variation of L&B prospecting intensity. The observed response curve of 1886 rare species richness to human population density may thus be an artefact of spatial heterogeneity in prospecting effort: many of rare species have been described in zones of intermediate human density, i.e. in the areas surrounding the main cities of the late XIXth century. As a result, this spatial

heterogeneity in 1886 sampling effort also strongly biases the analysis of rare species occurrences apparition from 1886 to 2001. However, prospecting intensity does not explain the spatial patterns in 2001 rare species richness and the spatial patterns of extinction and stability of rare species occurrences from 1886 to 2001.

Our results indeed indicated that the regional distribution of rare species occurrences respond significantly to combinations of physical and human factors. Spatial distribution of rare species occurrences is partly explained by topography and lithology, as documented in other regions (Heikkinen, 1998; Hill & Keddy, 1992; White & Miller, 1988; Wisser et al., 1998). In 1886 and 2001, rare species richness is particularly high in elevated zones, and on some particular bed-rocks such as Basalt, Limestone, and Shales / Quartzites. However, rare species richness has strongly increased from 2001 to 1886 in low-land areas (under 200m of elevation), on Alluvial-Colluvial deposition and on Dolomites-Chert Limestones. This discrepancy between 1886 and 2001 spatial pattern in rare species richness may partly result from varied regional history among the 2001 pool of rare species : a part of the 2001 rare species were not rare in 1886 but have been subject to a dramatic regression from that date, as for some species of the littoral habitats (e.g. *Othanthus maritimus* Hoffm.). The regional distribution of rare species occurrences is also sensitive to human activities and to their recent changes. In 2001, most of regionally rare species are thus present in zones of extensive agriculture, dominated by semi-natural open habitats, such as grasslands and heat-lands, and where percentage of cultivated area has been continuously regressing during the last decades. This pattern seems to be common in many agricultural or semi-natural landscapes (Fischer & Stöcklin, 1997; Luoto et al., 2002; Olsson et al., 2000).

The importance of biogeographical origin and range size of regionally rare species

In our study, rare species showed different responses to ecological and human factors according to their biogeographical origin and their total range size. The position of Mediterranean and Southern / Central Europe groups of rare species on the first OMI plane indicates that the different species of these two chorotypes can generally exploit a large array of ecological conditions. As a whole, Mediterranean and Southern / Central Europe rare species do not tend to occupy marginal conditions in the study region. On the contrary, Mountain and Eurosiberian rare species tend as a whole to occur in unusual ecological conditions, such as zones of high elevation and metamorphic substrates. The Mountain rare species are generally present in most of European mountains, and occur at their lower altitudinal range in the study region (e.g. *Minuartia laricifolia* (L.) Schinz & Thell., *Arabis alpina* L.). Most of Eurosiberian are widely distributed from the study region to Northern Europe: some of

them reach their range limits in the Mediterranean regions (e.g. *Actaea spicata* L., *Equisetum fluviatile* L.).

In the eventuality of environmental change in the region, the most marginal species are likely to first disappear from the region, depending whether the future conditions are unfavourable for these species. Global warming and/or land use change could thus constitute a threat for Mediterranean peripheral population of Eurosiberian species, and for the low elevation populations of Mountain species which persist in Mediterranean regions. Global change have already been observed as a factor of range shifting of some species (Hughes, 2000; Walther et al., 2002) and of range restriction of European Mountain species (Dirnbock et al., 2003). Mountain and Eurosiberian species should thus be at higher risk of extinction in the study region than any other chorotype of rare species.

However, spatial pattern of Mediterranean rare species richness is slightly different when the restricted endemic species are considered separately from other rare species. Endemic species richness responds positively to mean slope angle, and is higher in zones of low population density and where agriculture activity has been decreasing for the last century, i.e. with negative variation of cultivated area. The regional hotspots of endemic species are thus located in areas of heterogeneous topography, e.g.; with cliffs and rocky outcrops, and where human perturbations for urbanization and agriculture have been continuously low or decreasing. This result would indicate that endemic species constitute a group of regionally rare species with low sensitivity to land use change. Several physical or chemical characteristics of these habitats make them often so ecologically distinctive or even extreme that they support many restricted endemic species in many parts of the world (Baskin & Baskin, 1988; Matthews et al., 1993; McVaugh, 1943; Platt, 1951).

The importance of the regional history of rare species

The effects of human activities on the biodiversity are likely to be heterogeneously distributed, hence the importance to discriminate between the local conditions inducing rare species extinctions and the conditions providing the template for their long-term persistence. In the study region, some of the 2001 rare species were already rare in 1886 and remain stable between the two dates: these species are mainly located in zones of high elevation, high topographic heterogeneity (cliffs, rocky outcrops), and on basic bed-rocks such as basalt and limestones. Rare species occurrences thus had high persistence rates in zones where human activities may have been continuously low because of harsh topography and unfertile soils (e.g. *Antennaria dioica* (L.) Gaertner, *Aquilegia viscosa* Gouan, *Cheirolophus intybaceus* Dostal). This may explain the convergence observed between different regional studies to document higher rare species richness in

heterogeneous and inaccessible zones of semi-natural landscapes (Heikkinen, 1998; Luoto, 2000; White & Miller, 1988).

Rare species occurrences went extinct between 1886 and 2001 when present in zones where human populations, crops and animals breeding became the most densely distributed. Then, the most threatened regionally rare species are in the geographical zones where human perturbations have been intensifying, mainly because of increased tourism and agriculture productivity. As it has been shown at larger spatial scales, there is thus a spatial coincidence of zones of high biological value and zones of high economic value (Dobson et al., 2001). These results can appear very straightforward, but have the advantages of clearly demonstrating the strong deleterious effects that agriculture intensification and urbanisation can have on valuable component of biodiversity. This has been seldom demonstrated (Thompson & Jones, 1999), although the need for such evidence data has been claimed several times (Pimm, 2002).

Rarest species are generally removed from most studies of biogeography because their low number of occurrences prevent from analysing separately the spatial distribution of each rare species (Thuiller, 2003; Thuiller et al., 2003). However, regionally rare species often raise some conservation problems (Molina et al., 1999). By pooling the rarest species in our study region, our work may help to delimitate zones or species for which protection status would be the most relevant. Most of regionally rare species are present in high altitudes in zone of extensive agriculture, dominated by semi-natural open habitats, and where agriculture has been regressing between 1886 and 2001. Among these rare species, Eurosiberian and Mountain species are the most marginal in the region and seem to incur a high risk of regression in the Mediterranean under the current predictions of global change. The population of rare species are more likely to persist if they occur in zones where harsh topography and low fertility may limit accessibility and exploitation: many restricted endemic rare species seem to be in that case. On the contrary, regionally rare species are the most prone to extinction when occurring in zones of current urbanization and agriculture intensification. We hope these results to provide a solid scientific basis for a redistribution of conservation efforts among rare species of the Mediterranean regions.

LITERATURE CITED

Anon (1999) S-PLUS 2000 User's Guide Data Analysis Products Division, Seattle, WA.
 Baskin, J.M. & Baskin, C.C. (1988) Endemism in rock outcrop plant communities of unglaciated eastern United States : an evaluation of the roles of the edaphic, genetic and light factors. *Journal of Biogeography*, 15, 829-840.
 Choler, P. & Michalet, R. (2002) Niche differentiation and distribution of *Carex curvula* along a bioclimatic

gradient in the southwestern Alps. *Journal of Vegetation Science*, 13, 851-858.

Coste, H. (1900-1906) Flore descriptive et illustrée de la France et de la Corse et des contrées limitrophes Librairie scientifique et technique, Paris.

Daget, P. (1977) Le bioclimat méditerranéen: caractères généraux et modes de caractérisation. *Vegetatio*, 34, 1-20.

Debussche, M., Lepart, J., & Dervieux, A. (1999) Mediterranean landscape changes: evidence from old postcards. *Global Ecology and Biogeography Letters*, 8, 3-15.

Denelle, N., Bertrand, L., Granel de Solignac, L., Mazurek, H., & Schafer, P. (1995) Connaissance et exploration floristiques en Languedoc-Roussillon (France) : cartographie des points d'herborisation et répartition des Malvaceae pour l'Hérault. *Acta Botanica Gallica*, 142, 37-53.

Dirnbock, T., Dullinger, S., & Grabherr, G. (2003) A regional impact assessment of climate and land use change on alpine vegetation. *Journal of Biogeography*, 30, 401-417.

Dobson, A.P., Rodriguez, J.P., & Roberts, W.M. (2001) Synoptic tinkering: integrating strategies for large-scale conservation. *Ecological Applications*, 11, 1019-1026.

Dolédec, S., Chessel, D., & Gimaret-Carpentier, C. (2000) Niche separation in community analysis: a new method. *Ecology*, 81, 2914-2927.

Dugrand, R., ed. (1969) Atlas du Languedoc Roussillon. Berger-Levrault, Paris, France.

Dugrand, R. (1971) Atlas Régional du Languedoc-Roussillon Berger-Levrault, Paris.

Duncan, R.P. & Young, J.R. (2000) Determinants of plant extinction and rarity 145 years after European settlement of Auckland, New Zealand. *Ecology*, 81, 3048-3061.

Fiedler, P.L. (1986) Concepts of rarity in vascular plant species, with special reference to the genus *Calochortus* Pursh (Liliaceae). *Taxon*, 35, 502-518.

Fischer, M. & Stöcklin, J. (1997) Local extinctions of plants in remnants of extensively used calcareous grasslands 1950-1985. *Conservation Biology*, 11, 727-737.

Greuter, W. (1991) Botanical diversity, endemism, rarity, and extinction in the Mediterranean area: an analysis based on the published volumes of Med-Checklist. *Botanika Chronika*, 10, 63-79.

Greuter, W., Burdet, H.M., & Long, G. (1984-1989) Med-Checklist Conservatoire et Jardin botaniques de Genève, Genève.

Heikkinen, R.K. (1998) Can richness patterns of rarities be predicted from mesoscale atlas data? A case study of vascular plants in the Kevo Reserve. *Biological Conservation*, 83, 133-143.

Hill, N.M. & Keddy, P.A. (1992) Prediction of rarities from habitat variables: coastal plain plants on Nova Scotian lakeshores. *Ecology*, 73, 1852-1859.

Hughes, L. (2000) Biological consequences of global warming: is the signal already. *Trends in Ecology and Evolution*, 15, 56-61.

- Ihaca, R. & Gentleman, R. (1996) R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5, 299-314.
- Kerguelen, M. (1993) Index synonymique de la flore de France Museum National d'Histoire Naturelle, Paris.
- Le Floc'h, E., Long, G., Poissonet, J., & Godron, M. (1983). Végétation. In Atlas du Languedoc Roussillon (ed R. Dugrand). Berger-Levrault, Paris.
- Lepart, J. & Debussche, M. (1992). Human impact on landscape patterning: Mediterranean examples. In Landscape boundaries. Consequences for biotic diversity and ecological flows (eds A.J. Hansen & F. Di Castri). Springer-Verlag, New York.
- Loret, H. & Barrandon, A. (1886) Flore de Montpellier ou analyse descriptive des plantes vasculaires de l'Hérault, 2ème Edition LXXV, Montpellier.
- Luoto, M. (2000) Modelling of rare plant species richness by landscape variables in an agriculture area in Finland. *Plant Ecology*, 149, 157-168.
- Luoto, M., Toivonen, T., & Heikkinen, R.K. (2002) Prediction of total and rare plant species richness in agricultural landscapes from satellite images and topographic data. *Landscape Ecology*, 17, 195-217.
- Matthews, W.S., van Vyck, A.E., & Bredenkamp, G.J. (1993) Endemic flora of the north-eastern Transvaal Escarpment, South Africa. *Biological Conservation*, 63, 83-94.
- McVaugh, R. (1943) The vegetation of the granitic flat-rocks of the southeastern United States. *Ecological Monographs*, 13, 121-166.
- Molina, J., Mathez, J., Debussche, M., Michaud, H., & Henry, J.P. (1999) Méthode pour établir une liste régionale d'espèces protégées. Application à la flore du Languedoc-Roussillon. In Les plantes menacées de France (ed J.Y. Le Souëff), Vol. 19, pp. 399-420. Bull.Soc.Bot.Centre-Ouest, Brest.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., B., D.F.G.A., & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, 403, 853-858.
- Olsson, E.G.A., Austrhaim, G., & Grenne, S.N. (2000) Landscape change patterns in mountains, land use and environmental diversity, Mid-Norway 1960-1993. *Landscape Ecology*, 15, 155-170.
- Parmesan, C. & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37-42.
- Pimm, S.L. (1996) Lessons from a kill. *Biodiversity and Conservation*, 5, 1059-1067.
- Pimm, S.L. (2002) The Dodo went extinct (and other ecological myths). *Annals of the Missouri Botanical Garden*, 89, 190-198.
- Platt, R.B. (1951) An ecological study of the Mid-Appalachian shale barrens and of the plants endemic to them. *Ecological Monographs*, 21, 269-300.
- Pons, A. & Quézel, P. (1985). The history of the flora and vegetation and past and present human disturbance in the Mediterranean region. In Plant conservation in the Mediterranean area (ed C. Gomez-Campo), Vol. 7, pp. 9-24. Geobotany, Dordrecht, The Netherlands.
- Prendergast, J.R., Quinn, R.M., Lawton, J.H., Eversham, B.C., & Gibbons, D.W. (1993a) Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature*, 365, 335-337.
- Prendergast, J.R., Wood, S.N., Lawton, J.H., & Eversham, B.C. (1993b) Correcting for variation in recording effort in analyses of diversity hotspots. *Biodiversity letters*, 1, 39-53.
- Quézel, P. (1985). Definition of the Mediterranean region and the origin of its flora. In Plant conservation in the Mediterranean area (ed C. Gomez-Campo), Vol. 7, pp. 9-24. Geobotany, Dordrecht, The Netherlands.
- Sala, O.E., Chapin, F.S.I., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M., Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M., & Wall, D.H. (2000) Global biodiversity scenarios for the year 2000. *Science*, 287, 1770-1774.
- Thioulouse, J., Chessel, D., Dolédec, S., & Olivier, J.M. (1997) ADE4: a multivariate analysis and graphical display software. *Statistics and Computing*, 7, 75-83.
- Thompson, K. & Jones, A. (1999) Human population density and prediction of local plant extinction in Britain. *Conservation Biology*, 13, 185-189.
- 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., Vayreda, J., Pino, J., Sabate, S., Lavorel, S., & Gracia, C. (2003) Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology and Biogeography*, 12, 313-325.
- Tutin, T.G., Burges, N.A., Chater, A.O., Edmondson, J.R., Heywood, V.H., Moore, D.M., Valentine, D.H., Walters, S.M., & Webb, D.A. (1964-1993) *Flora Europaea* Cambridge University Press, Cambridge.
- Venable, W.N. & Ripley, B.D. (2002) *Modern applied statistic with S* Springer, New York.
- Venables, W.N. & Ripley, B.D. (2002) *Modern applied statistic with S* Springer.
- Vitousek, P.M., Mooney, H.A., Lubchenco, J., & Melilo, J.M. (1997) Human domination of earth's ecosystems. *Science*, 277, 494-499.
- Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.-M., Hoegh-Guldberg, O., & Bairlein, F. (2002) Ecological responses to recent climate change. *Nature*, 416, 389-395.
- White, P.S. & Miller, R.I. (1988) Topographic models of vascular plant richness in the southern appalachian high peaks. *Journal of Ecology*, 76, 192-199.
- Wiser, S.K., Peet, R.K., & White, P.S. (1998) Prediction of rare-plant occurrence: a southern appalachian example. *Ecological Applications*, 8, 909-920.

Table 1 : Description of explanatory variables used to model rare species richness in Hérault Department. See text for further explanations on the variables.

Variables	Code	Derivation / calculation	Source
Prospection intensity for 1886 survey (%)	Prop_Intens	Percentage of proposed communes per arrondissement	Denelle <i>et al.</i> 1995
Lithology	Lithology	9 classes (see text)	Dugrand 1969
Mean slope angle (°)	Slope_Mean	Derived with Arcview 3D analyst	Digital elevation model (50x50m)
Mean altitude (m)	Alti_Mean	Derived with Arcview 3D analyst	Digital elevation model (50x50m)
Population density in 1881 (hab nb / km ²)	PopDens1881	Inhabitants number per commune / commune area	1881 national census
Population density in 1954 (hab nb / km ²)	PopDens1954	Inhabitants number per commune / commune area	1954 national census
Population density in 1999 (hab nb / km ²)	PopDens1999	Inhabitants number per commune / commune area	1999 national census
Population density variation from 1954 to 1999 (%)	VarPop1954	100 * (PopDens1999 - PopDens1954) / PopDens1954	1954 and 1999 national census
Population density variation from 1881 to 1999 (%)	VarPop1881	100 * (PopDens1999 - PopDens1881) / PopDens1881	1881 and 1999 national census
Grassland and heathland area (%)	PctSth2000	100 * Grassland and heathland area / Commune area	2000 agricultural census
Total cultivated area (%)	PctSau2000	100 * Total cultivated area / Commune area	2000 agricultural census
Total livestock density (head number / km ²)	DnsChp2000	Ovine + caprine livestock / commune area	2000 agricultural census
Grassland / heathland area variation from 1970 to 2000 (%)	VarSth1970	100 * (PctSth2000 - PctSth1970) / PctSth1970	1970 and 2000 agricultural censuses
Total cultivated area variation from 1970 to 2000 (%)	VarSau1970	100 * (PctSau2000 - PctSau1970) / PctSau1970	1970 and 2000 agricultural censuses
Total cheptel density variation from 1970 to 2000 (%)	VarChp1970	100 * (DnsChp2000 - DnsChp1970) / DnsChp1970	1970 and 2000 agricultural censuses

Table 2 : Results of stepwise selection for the 4 most important variables in GLM for rare species richness in 2001 (Rare sp. 2001), rare species richness in 1886 (Rare sp. 1886), narrow endemic species richness in 2001 (Endemic sp. 2001), number of rare species extinctions from 1886 to 2001 (Rare sp. extinction 1886-2001), occurrence of rare species with stable dynamics from 1886 to 2001 (Rare sp. stability 1886-2001). The four selected explicative variables are listed in order of decreasing deviance explained. Results are presented for GLMp, i.e. GLM allowing for 2nd order polynomial terms (e.g. pol(Slope_Mean) being a 2nd order polynome for Slope_Mean effect) and for GLMs, i.e. GLM with linear terms only. The last column gives the Chi² statistic and its associated degrees of freedom (in subscript) for the log-likelihood ratio test, which compares the two GLM, with ^{ns} : not significant, **: $P < 0.01$, ***: $P < 0.001$. The best model (with the highest explained deviance) is written in boldface.

	GLMp	GLMs	Chi ² df
Rare sp. 1886	pol(Alti_mean), pol(PopDens1881), Prop_Intens, Lithology	Lithology, Prop_Intens, PopDens1881, Alti_mean	21.9 ₄ ***
Rare sp. 2001	pol(Alti_mean), Lithology, pol(PctSth2000), VarSau1970	Lithology, Alti_mean, VarSau1970, VarChp1970	78.5 ₅ ***
Endemic sp. 1886	pol(Slope_Mean), pol(Prop_Intens)	Slope_Mean, Prop_Intens	4.65 ₁ *
Endemic sp. 2001	Slope_Mean, pol(PctSth2000), VarSau1970, PopDens1999	Slope_Mean, VarSau1970, VarChp1970	17.9 ₄ **
Rare sp. extinction 1886-2001	pol(VarPop1881), VarChp1970, pol(VarPop1954), VarSau1970	VarPop1881, VarChp1970, VarSau1970, Lithology	4.03 ₅ ^{ns}
Rare sp. stability 1886-2001	Lithology, pol(Alti_Mean), pol(Slope_Mean), pol(PctSau2000)	Lithology, Slope_Mean, Alti_Mean, PctSth2000	14.2 ₃ **



Figure 1: Geographical location of Hérault department, the study area (black line) in the Méditerranéan Basin. Names of main cities, lake and lagoons are given.

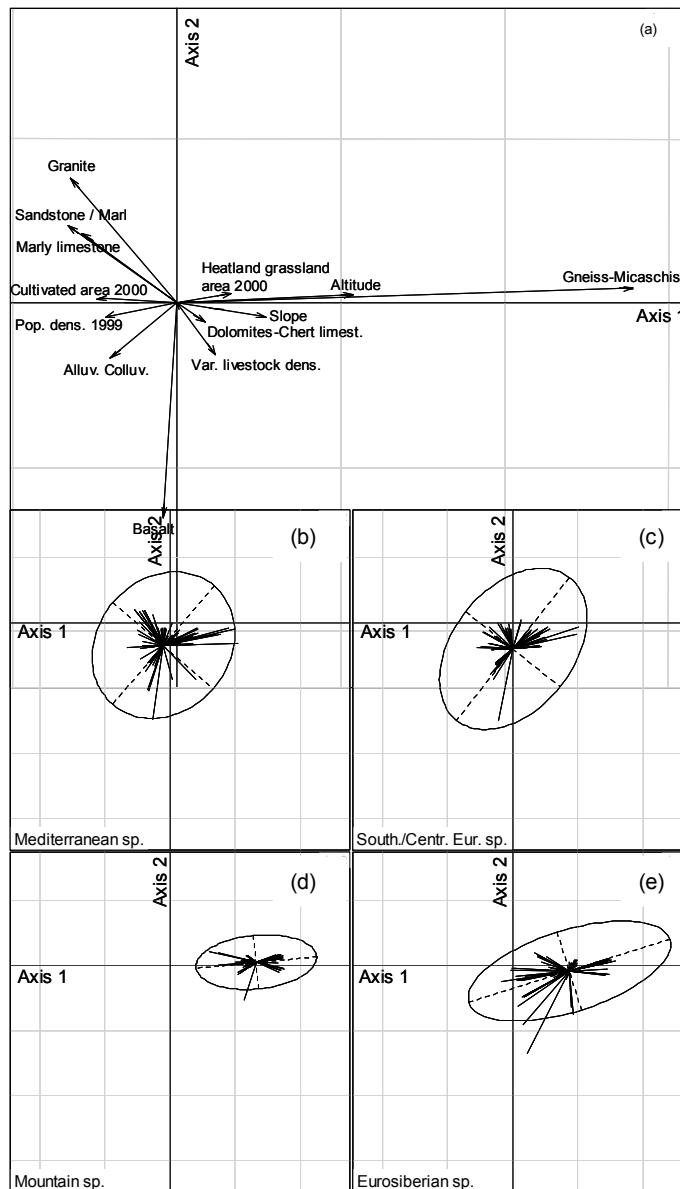


Figure 2: Contribution of environmental, human and land use variables to the two first axes produced by the OMI analysis (a), with population density in 1999 (Pop.dens.1999), alluvial-colluvial deposit (alluv.colluv.), Variation of livestock density from 1970 to 2000 (Var.livestock dens.). The ecological position of the four chorological group of rare species are represented on the first plane of OMI analysis with Mediterranean / Sub-Mediterranean rare species (b), Western and Central Europe rare species (c), Mountain rare species (d), Eurosiberian rare species (e).

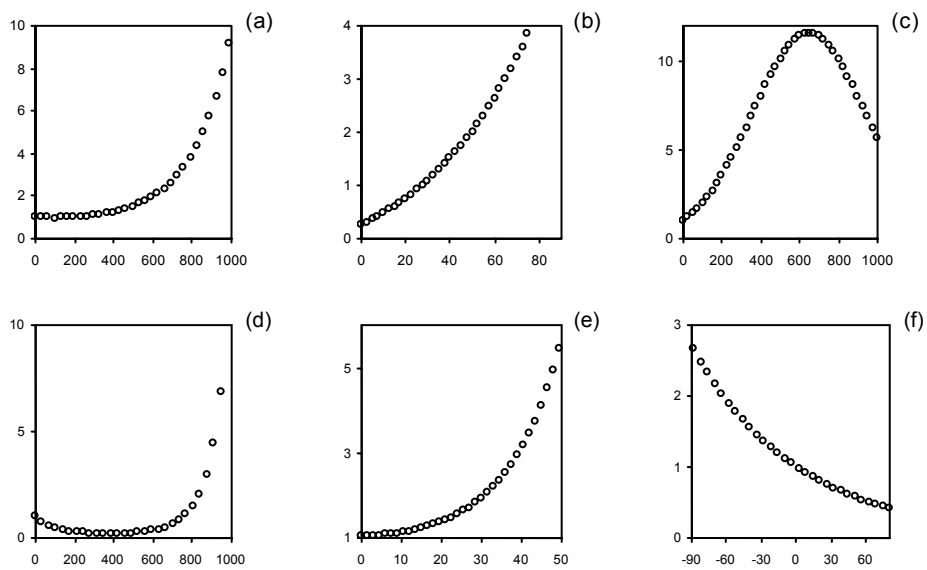


Figure 3: Fitted values of rare species richness in 1886 as a function of mean altitude (a), L&B prospection intensity in 1886 (b), population density in 1881 (c). Fitted values of rare species richness in 2001 as a function of mean altitude (d), grassland and heathland area percentage (e), cultivated area variation from 1970 to 2000 (f). Fitted values of rare species richness were calculated by exponential transformation of estimates generated by GLM.

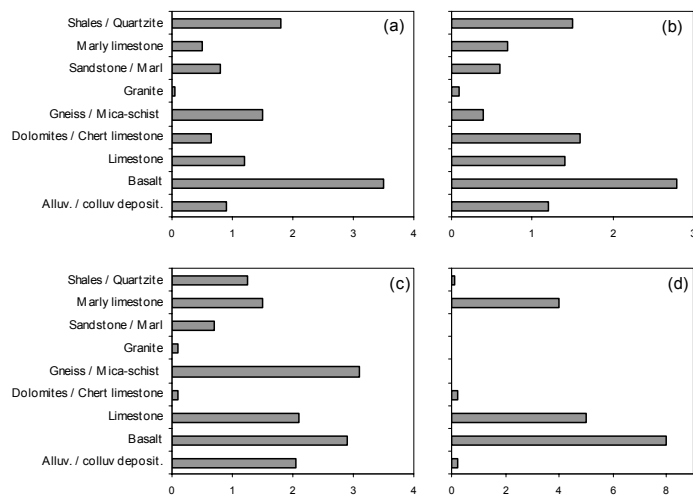


Figure 4: Fitted values of rare species richness in 1886 (a), rare species richness in 2001 (b), number of rare species extinctions from 1886 to 2001 (c), and number of rare species with stable dynamics from 1886 to 2001 (d) between different lithology classes.

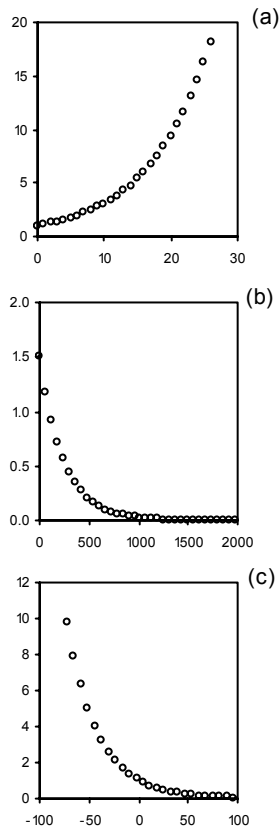


Figure 5: Fitted values of endemic species richness in 2001 as a function of mean slope angle (a), population density in 1999 (b), variation of cultivated area from 1970 to 2000 (c). Fitted values of rare species richness were calculated by exponential transformation of estimates generated by GLM.

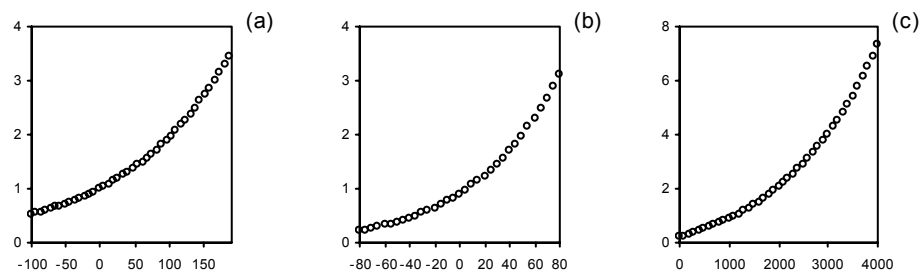


Figure 6: Fitted values for number of rare species extinctions from 1886 to 2001 as a function of variation of livestock density between 1970 and 2000 (a), variation of cultivated area from 1970 to 2000 (b), Variation of population density from 1881 to 1999 (c). Fitted values of rare species richness were calculated by exponential transformation of estimates generated by GLM.

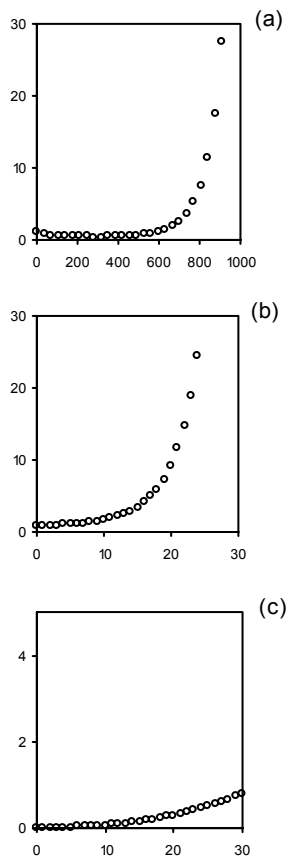


Figure 7: Fitted values for occurrence of rare species with stable dynamics from 1886 to 2001 as a function of mean altitude (a) and mean slope angle (b). Fitted values of rare species richness were calculated by exponential transformation of estimates generated by GLM.

ANNEXE 4.2

Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* species in the Cape Floristic Region

Thuiller W., Lavorel S., Midgley G.F., Lavergne S., Rebelo A.G.

Ecology (2004), *sous presse*

Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* taxa in the Cape Floristic Region

Wilfried Thuiller¹, Sandra Lavorel², Guy Midgley^{3,4}, Sébastien Lavergne¹ and Tony Rebelo³

¹Centre d'Ecologie Fonctionnelle et Evolutive, CNRS. 1919 route de Mende, 34293 Montpellier Cedex 5, France

²Laboratoire d'Ecologie Alpine. CNRS. Université J. Fournier. BP 53X, 38041 Grenoble Cedex 9, France

³Climate Change Research Group, Kirstenbosch Research Center, National Botanical Institute, P/Bag x7, Claremont 7735, Cape Town, South Africa.

⁴Center for Applied Biodiversity Science, Conservation International, 1919 M St., NW, Washington, D.C. 20036

Corresponding author: Wilfried Thuiller

E-mail: wilfried.thuiller@cefe.cnrs-mop.fr

ABSTRACT

Convergence between species niches and biological traits was investigated for 88 *Leucadendron* taxa in Cape Floristic Region. Firstly, niche separation analysis was performed to relate species' niche positions/breadths with bioclimatic gradients. These gradients of aridity, seasonality of water availability, heat and cold stress explained almost all variation in niche distributions. Species present in zones of extreme aridity or temperature exhibited narrower niche breadths than species situated in moderate sites, suggesting that stress-tolerant species do not occupy broad environmental ranges. Secondly, species niche positions were related to selected biological traits. Species of arid sites had significantly lower blade areas than did species of moist sites, confirming a functional trade-off between stress tolerance and productivity for leaf design. Species

dispersal mode was correlated to species niche positions on the aridity gradient, suggesting allometrically determined correlations between leaf design and the design of reproductive structures. Species niche positions were also correlated with flowering traits, with species initiating flowering in winter found under Mediterranean climate influence and species initiating flowering in spring in sites with greater summer rainfall input. By interrelating species niche positions on bioclimatic gradients with selected biological traits, we explore a novel biogeographical approach to understanding species distributions.

KEYWORDS: Climatic niches, Gradient analysis, Niche breadth, Niche position, Outlying mean index analysis (OMI), morphological traits, dispersal mode, flowering phenology.

INTRODUCTION

Hutchinson (1957), defining the concept of ecological niche, suggested that the niche could be modeled as an imaginary space with many dimensions, in which each dimension or axis represents the range of some environmental variable required by the species. Thus, the niche of a plant might include the range of temperatures that it can tolerate, the intensity of light required for photosynthesis, specific humidity regimes, or minimum quantities of essential soil nutrients for uptake (Huston 1994). A useful extension of the niche concept is the distinction between fundamental and realized niches (Hutchinson 1957). The fundamental niche of a species includes the total range of environmental conditions that are suitable for existence without the influence of inter-specific competition, predation, dispersal limitation and natural or human disturbances (Pulliam 2000). The

realized niche describes that part of the fundamental niche actually occupied by the species under these constraints and defines the spatial distribution of species in a community and in a study area (Austin and Smith 1989).

Correlative methods used to predict current or future spatial distributions of plant species are based on the concept of realized niche (Thuiller 2003, Thuiller et al. 2003). A range of methods has been used to describe and separate ecological species niches in an environmental space, including canonical correspondence analysis (ter Braak 1986) or redundancy analysis (ter Braak 1987). Several studies have related species niches to bioclimatic conditions and/or to geological and land use conditions to describe and predict environmental niches positions and tolerances, or to forecast the effects of climatic

change on the dynamics and patterns of species distributions (Malanson et al. 1992).

However little concern has been applied to establish a link between physiological, morphological and reproductive attributes of species and their niche positions and/or tolerance (but see Niinemets 2001). Thus, it remains a challenge to choose ecologically meaningful functional attributes to understand and predict species distribution (Grime 1974, 1979, Weiher et al. 1999). Several studies have attempted to select such traits (e.g. specific leaf area or plant height) *a priori* to define plant functional types (PFT) and forecast impacts of global climatic change (Díaz and Cabido 1997, Westoby 1998). Other authors have identified biological traits directly influenced by climatic factors by measuring the degree of correlation between corresponding climatic gradients (e.g. temperatures or precipitations) and biological traits of species distributed along them (Parkhurst and Loucks 1972, Fonseca et al. 2000). When few species are studied, 'hard' plant physiological traits can be measured accurately and comprehensively (e.g. growth rate or individual biomass). But when large numbers of species are examined, more easily measured 'soft' traits, often morphological and some times reproductive, must be used (Hodgson et al. 1999). Hypotheses must therefore be derived, based on these studies and the ecophysiological literature, to select such soft traits (Grime et al. 1997, Hodgson et al. 1999). Leaf parameters (e.g. leaf area, specific leaf area and leaf thickness) and plant height have appeared correlated with climatic conditions (Díaz et al. 1998). Other traits such as dispersal modes (Díaz et al. 1998), seed reserve mass and specific root length (Wright and Westoby 1999), flowering dates and duration (Hodgson 1999) are thought to relate to species distributions by determining their colonization abilities. Analyses of this nature have been made possible by the recent availability of databases of morphological and reproductive traits to deal with significant numbers of species and large geographical scales (Grime et al. 1997, Reich et al. 1999).

The aim of this paper is to propose an original approach to examine relationships between species biological traits and their distributions along bioclimatic gradients over large geographical scales. First we identified the main climatic gradients separating species niches. A multivariate analysis is the best way to perform such an analysis, as it allows an overall view of the species assemblage. We used a recently described multivariate method, the outlying mean index analysis (OMI analysis), that allows us to separate species niches and to measure the distance between the mean habitat conditions used by each species and the mean habitat conditions of the study area (Doledec et al. 2000). The second step of our approach, which to our knowledge is entirely original, consisted in correlating species niche

breadths (variability of used habitats) and mean positions along climatic gradients obtained with the OMI with their morphological (height and leaf area) and reproductive traits (dispersal mode and flowering phenology).

Multi-species comparative studies such as this can be accused of phylogenetic mismatching – i.e. comparing species traits among species with very different evolutionary histories. One way to control for phylogenetic history is to apply this analysis to related species, preferably within a single genus, comprising species with a wide range of traits and occupying a broad geographical area. The genus *Leucadendron* provides just such an opportunity, as it is speciose, endemic to a constrained geographic region in the Cape Floristic Region in South Africa, and has known monophyletic origin. The genus *Leucadendron* comprises short to tall evergreen shrubs with a variety of ecological requirements and contains both widespread and rare species.

Our analysis addressed the following questions:

- 1) What are the climatic gradients separating niches among *Leucadendron* species of the Cape Region?
- 2) What are the distributional amplitudes of different species on these main gradients?
- 3) Can species responses to these climatic gradients be related to biological traits, allowing for a first identification of climatic response groups?

DATA AND METHODS

Data set:

Location: the study area is situated in the Cape Floristic Region (CFR hereafter), in the southwestern corner of southern Africa. This region is made up of lowland coastal sandy plains and adjacent inland mountains of the Cape Fold Belt, with a sandstone dominated geology. The region is a known biodiversity hotspot, and is covered by Fynbos vegetation which can be described as a sclerophyll scrubland dominated by small leaved shrubs (Ericaceae), aphyllous grass-like forms (Restionaceae), and broad leaved shrubs (Proteaceae) (Goldblatt and Manning 2000). Fynbos is fire prone vegetation, and burns with a frequency of between 5 and 25 years, but generally between 10 and 20 years.

The genus *Leucadendron* (Proteaceae) is endemic to the CFR, and comprises 88 taxa (species and distinct subspecies) found throughout the region (Williams 1972). Species in this genus are evergreen low to mid-high shrubs (between 0.5 and 3m tall), and are dioecious, with females in half the species retaining seeds in serotinous cones until they are released after occasional fires, the remainder cached by ants and rodents (Rebelo 2001). The genus has many species which are killed by fire and obligatory dependent on seed regeneration

(seeders), and a few species which are able to resprout after fire, and thus not obligatory dependent on seed regeneration (sprouters). Seed dispersal in most species in this genus is predominantly by wind dispersal, and this mode is common in several Proteaceae (Bond 1988), but in some species either ants or rodents disperse seeds. The spatial distribution of species was recorded in a massive effort to atlas the family Proteaceae in southern Africa (Rebello 1992). Species presences were noted at more than 30 000 sites throughout the southern Cape region, and are available in GIS format (Rebello 2001). Species morphological traits were extracted from a related database (Rebello 2001).

Climatic variables: The climatic data set used for this study contained 10 climatically derived variables considered critical to plant physiological function and survival. Variables used (see also Midgley et al. 2003) were mean annual temperature (MAT), mean minimum temperature of the coldest month (MMTC), mean maximum temperature of the warmest month (MMTW), heat units (annual sum of daily temperatures ($^{\circ}\text{C}$) exceeding 18°C ; Htunt), mean annual precipitation (MAP), summer (Sumprec) and winter precipitation (Winprec), annual potential evaporation (Apann), winter soil moisture days (Smdwin), and summer soil moisture days (Smdsum). The latter parameters refer to the number of days for which soil moisture was sufficient and air temperature appropriate for plant growth.

Niche separation.

A number of multivariate methods are available to determine and separate species niches (Green 1971, Austin et al. 1984). The best-known technique is correspondence analysis (CA: Hill 1974), which is particularly well-adapted for investigating habitat amplitudes of species (Chessel et al. 1982). CA has supplied a basis for more complex approaches to relate the distribution of species in a study area to environmental factors. Canonical correspondence analysis (CCA) was originally developed to separate species niches along environmental gradients (ter Braak 1986) under the assumption of unimodal response curves. Similarly, redundancy analysis (RDA: ter Braak 1987) has been used for analysing linear relationships between species and gradient and is particularly well adapted to situations with one strong limiting factor (Doledec et al. 2000).

Recently, Doledec et al. (2000) proposed a new multivariate analysis (implemented in ADE4 software: Thioulouse et al. 1997) that makes no assumption about the shape of species response curves to the environment and, unlike CCA and RDA, gives equal weight to species-rich and species-poor sites. This technique appeared as particularly suitable for our purpose. The result of this analysis describes the mean position of the species in the environmental space, which represents a measure of the distance between the mean habitat

conditions used by the species and the mean habitat conditions of the study area. It measures the propensity of the species to select a specialized environment. Therefore, following the main gradients (or axes), we could analyze the variability of habitats used by each species ('niche breadth')

Niche separation and biological traits

We analyzed the relationship between species niche position on the main gradients and the biological traits selected as relevant to response to climate. These traits included quantitative (height and leaf area) and categorical traits (dispersal modes, season of beginning and end of flowering).

To analyze relationships between quantitative traits and the mean position of species on gradients, we used Least Trimmed Squares (LTS) robust regressions, which are useful for fitting linear relationships when the data are not Gaussian and contain significant outliers (Rousseeuw and Leroy 1987). Like a usual linear regression, LTS provides an R-squared test value (called robust- R^2) that evaluates variance explained by the independent variables. LTS regressions were performed with the software Splus 2000 (Anon 1999).

We used non-parametrical tests (Kruskall-Wallis and Wilcoxon tests) to analyze the relationships between species niche position on bioclimatic gradients and categorical traits, for which the homoscedasticity (Levene test: $p > 0.05$) and Gaussian conditions were violated so that ANOVA could not be used. Kruskal-Wallis tests were performed to detect significant differences among traits attributes according to species niche positions. Wilcoxon tests were used to make pairwise comparisons across attributes.

RESULTS

Niche separation

Although not presented in detail, CCA and RDA were also performed at the same time as the OMI. There was evidence that OMI was more suitable than CCA or RDA to separate climatic species niches explaining 94% of the total variability (Table 1). Ordination diagrams on the first two axes of the OMI analysis separated species niches according to two main gradients. Aridity, represented by a precipitation-evapotranspiration gradient, was the strongest gradient affecting species distributions (Table 2 and Figure 1A). This factor did not have the same impact on all *Leucadendron* species (Figure 2A). Species occurring in arid sites (*L. foedum* and *L. remotum*) had narrower niches (specialist species), than species present in more humid zones (generalist species) like *L. conicum* and *L. gandogeri*. However, some species found at sites in the middle of the gradient also had narrow niches. Hence, the aridity gradient separated out water-stress tolerant species with a narrow climatic niche from generalist species

occurring in more mesic sites. The second climatic gradient (Figure 1B) appeared to represent seasonality of water availability (Table 2). This gradient was interpreted as a gradient from mediterranean (high water availability in winter) to sub-tropical (low water availability in winter but high in summer) influence. As occurred in the case of the aridity gradient, there were differences in terms of niche breadth across the gradient (Figure 2B). Species situated in mediterranean-climate sites (e.g. *L. daphnoides* and *L. macowanii*) seemed to be more specialized than species found in more sub-tropical areas (e.g. *L. rourkei* and *L. album*). The ordination diagram therefore separated species climatic niches into four zones: mediterranean humid, mediterranean dry, subtropical humid and subtropical dry (Figure 3) corresponding respectively to the southwestern Cape mountains, western Cape Mountains and coastal lowlands, southern Cape coastal lowlands and eastern Cape interior.

The ordination diagram with the first and third axes displayed a third gradient driven by temperature (minimum temperature of the coldest month and mean annual temperature variables) (Figure 1B), separating species tolerant according to their tolerance of warm temperatures. As observed for the first two axes, this gradient also separated specialist and generalist species, and species tolerant to high temperatures or not (Figure 2C). Species present in sites with high mean temperature had narrower niches (e.g. *L. corymbosum* and *L. lanigerium* var. *leavigatum*) than species occurring elsewhere (*L. arcuatum* and *L. pubescens*).

Niche positions and biological traits

Relation between niche positions and dispersal modes: Significant differences between positions on axis 2 indicated that species dispersed by ants were preferentially found in subtropical sites (low soil moisture in winter: Figure 4B) compared to species dispersed by wind, present on the entire gradient but preferentially in mediterranean-climate zones (Wilcoxon test: $P=0.008$). The other significant difference was between species dispersed by rodents and wind (Wilcoxon test: $P=0.031$). Species dispersed by rodents were mostly present in the middle of the gradient, in climates typical of the mountains and lowlands of the southern Cape (Figure 4B). Relationships between dispersal modes and the mean position of species on of first and third axes representing aridity and temperature respectively were on the other hand not significant (Wilcoxon tests $P>0.05$) (Figure 4A, C).

Relation between niche positions and morphological traits: There was no relationship between the mean position of species on gradients and plant height. Leaf area was correlated with the mean position of species on the aridity gradient (Figure 5). Species present in arid sites had smaller individual leaf blade areas than species present in more humid habitats. Likewise, species

occurring at sites with high potential evapotranspiration had small leaves, while species occurring in sites with high annual precipitation had larger leaves.

There was no relationship between leaf area and the mean position of species on both OMI axis 2 (seasonality of water availability) and OMI axis 3 (temperature) (robust- $R^2 = 0.8\%$ and $R^2 = 1.6\%$ respectively).

Relationships between dispersal modes and morphological traits: To further clarify the relationships between dispersal mode and species mean position on the three main gradients, we analyzed relationships between morphological traits (leaf area and height) and dispersal mode (Figure 6). Species dispersed by ants or rodents were significantly shorter than wind-dispersed species (Wilcoxon test: $P = 0.007$ and $P = 0.02$ respectively) (Figure 6A). The apparent difference between species dispersed by ants and rodents was not significant (Wilcoxon test: $P > 0.05$).

Species dispersed by ants had smaller leaf area than species dispersed by rodents and wind (Wilcoxon test: $P < 0.001$ and $P < 0.001$). Species dispersed by rodents had greater leaf area than species dispersed by wind (Wilcoxon test: $P < 0.001$) (Figure 6B).

Relationships between niche positions and flowering phenology:

Dates of onset of flowering differed significantly according to the positions of species on the gradient of seasonality of water availability gradient (Kruskall-Wallis test: $P<0.001$) (Figure 7). Species occurring in mediterranean-climate sites tended to flower more frequently in winter, while in subtropical sites, species flowered preferentially in spring (Wilcoxon test: $P = 0.000$). Similarly, the date of the onset of flowering changed significantly along the aridity gradient (Wilcoxon test: $P = 0.023$) (Figure 7A). Thus the beginning of flowering appeared highly dependent on the type of eco-region in which a species is present. Species found under a mediterranean climate initiate flowering mainly in winter, while species found in the humid subtropical zone flower in spring, and species occurring under dry subtropical climate flower mainly in autumn.

Dates of end of flowering differed significantly according to the positions of species on the gradient of aridity (Kruskall-Wallis test: $P = 0.0276$) (Figure 7D). Similarly, the date of the end of flowering changed significantly along the seasonality of water availability gradient (Wilcoxon test: $P = 0.0259$) (Figure 7E). Species occurring in humid subtropical regions ended their flowering in spring rather than in winter and therefore had a shorter flowering period than species from other climatic areas.

DISCUSSION

Niche separation: The OMI analysis provided a powerful framework to separate niches of 88 *Leucadendron* species in the Cape Floristic region. We identified three main gradients governing the distributions of *Leucadendron* species. The gradients of aridity and precipitation seasonality defined four eco-regions: arid and humid Mediterranean, arid and humid subtropical zones (roughly coincident with Köppen and Holdridge classifications for this region as mapped by Schulze 1973). A third temperature gradient separated species tolerant to hot from those tolerant to cooler conditions. On each of these gradients the *Leucadendron* genus included both generalists, with a wide distribution along a given gradient, and specialists of specific habitat positions, usually towards the more 'stressful' part of the gradient (dry, hot or with a winter rainfall regime). Specialists may be constrained to these conditions by competition or represent true stress-tolerant species.

Relationships among niches and soft plant traits:

Based on the theory of natural selection we would expect species with similar niches to share similar biological characteristics. Species in more arid sites were indeed found to have smaller leaves than species occurring in more humid sites, a pattern commonly observed in a range of floras (Parkhurst and Loucks 1972, Givnish 1987, Fonseca et al. 2000). Physiologically, species occurring in arid environment develop strategies to limit the loss of water by reducing the exchange area with air. In terms of ecological strategies, species tolerant of water-stress were mainly found in specialized habitats of arid zones, which possibly reflected a poorer competitive ability than generalists in more humid habitats. Inversely, species with larger leaves had a more generalist behavior, with widespread distributions and high habitat occupancy. This type of distribution may reflect a high competitiveness across a range of habitats, based on which, and in accordance with previous authors (Grime et al. 1997, Westoby et al. 2002) individual leaf area may be interpreted as a trait associated with high competitiveness.

On the other hand, contrary to what might have been expected from plant strategy schemes (Grime 1979, Westoby 1998) and previous regional studies (Díaz and Cabido 1997), plant response to climate was independent of plant height. Considering that the analysis was conducted within a single genus representing a single life form it may be worth considering a more detailed measure of plant aboveground biomass than plant height (e.g. canopy volume, if not biomass itself). We can also not exclude confounding phylogenetic effects that may hide a general relationship between plant height and climatic distribution (Ackerly et al. 2000), though the likelihood is low due to the group's monophyly. Finally, our analysis did not consider a complicating factor,

and in particular fire, which represents a major disturbance in the study region, but for which no regional data set was available. Should plant height be related to response to disturbance (Grime 1979) the expected response (smaller plants in more disturbed habitats) would run counter to that expected for climate (taller plants in more mesic habitats) because, over the climatic range of interest, the importance of fire is related to fuel load, which increases with decreasing aridity. Taken together, this combination of factors and expected responses would yield a globally neutral response.

Changes in leaf area along the aridity gradient may be the main cause of the relationship observed between dispersal mode and climate. Indeed, leaf area increased from species dispersed by ants to species dispersed by wind and species dispersed by rodents. Cornelissen (1999) demonstrated for 58 British semi-woody species that small seed mass and leaf area were found among species from stress-prone and early successional habitats, while large seeds and leaves were found among competitive plants of later successional stages. In our study, this pattern may be explained by climatic and developmental constraints. Water stress decreases leaf size, thereby narrowing cone size and total seed mass by an allometric effect (Bond and Midgley 1988). This relation implies strong environmental selection on dispersal adaptations in this genus which has evolved a remarkable range of dispersal mechanisms. In particular, dispersal by rodents (which is possibly unique to this genus in the Fynbos Biome) requires plants with large biomass that can afford a high absolute energy allocation to develop large seeds. Such interpretations would need to be tested by measurements of plant and seed biomass across habitat types. Finally, our results seem to support the hypothesis that species growing at the edge of environmental gradients ('stress-tolerators') are slow growing, assuming that small leaves - and associated low dispersal ability (by ants) - implies slow growth (Grime et al. 1997).

Flowering phenology of *Leucadendron* species also varied with species niche positions along climatic gradients of the Cape region. We showed that species growing in similar eco-regions (humid and arid mediterranean and humid and arid subtropical eco-region) have developed similar phenologies. Consistent with Europe, species beginning to flower in winter are mainly found in the Mediterranean-climate zone. Our analysis indicated that in South Africa coexisting *Leucadendron* species, share attributes that give similar flowering phenology.

This study shows that to define reliable functional groups sharing similar traits and response to environment, a preliminary biogeographical study is needed to separate ecological niches, followed by the correlation of these to functional attributes. The biological traits used for this study were dependent on

data availability, but additional analyses using traits with better-established functional significance could further the study of patterns of adaptive evolution under climatic constraints. Westoby (1998) proposed the use of leaf mass per area (LMA), seed mass and specific root length (SRL). In addition, other traits relevant to climate response could be used, such as leaf dry matter content (LDMC) (Niinemets 2001), or seedling relative growth rate (RGR) and seed reserve mass (Wright and Westoby 1999). For instance, average LMA is well known to be higher at low rainfall, owing to thicker leaves, denser tissue, or both Westoby et al. 2002. Further useful insights into leaf design in this group would derive from study of the geographic variation in LMA.

Relevance for global change studies:

Rapid changes of climate and land use could have unexpected consequences for biodiversity (Chapin III et al. 2000). Because of the enormous complexity of individual species and populations, an alternative to assess the effect of global change on diversity is to summarize species information into a relatively small number of general recurrent patterns: functional groups (Woodward and Cramer 1996, Smith et al. 1997). However, to date few studies have directly analyzed the response of functional groups, determined from direct distribution analyses and not *a priori* defined, to global change. Instead global biome model to assess impacts of future climates changes on potential natural vegetation patterns have used plant functional types based on plant life forms and phenology (e.g. Prentice et al. 1992, Bachelet et al. 2001).

A more comprehensive framework would involve defining climate response groups based on responses of species distributions to climatic variables and on their functional attributes, as presented in this study. This paper presents one possible approach that simultaneously links distribution data, environmental variables and functional traits (Doledec et al. 1996, Legendre et al. 1997). Niche-based models could then be used to assess the impact of climate change on these groups by predicting their current climatic distributions and projecting future climatic envelopes under scenarios of climate change, as is commonly done for species distributions (e.g. Iverson and Prasad 1998, Thuiller 2003).

ACKNOWLEDGMENTS

This research was supported by the Ministère des Affaires Etrangères et Enseignement Supérieur of France and the National Research Foundation of South Africa. This paper also contributes to the Montpellier – Barcelona European Laboratory (LEA, CNRS) "Mediterranean Ecosystems in a Changing World". The first author is funded by the European Union's Advanced Terrestrial Ecosystem Analysis and Modelling project

(EVK2-CT-2000-00075). GFM acknowledges the support of the Centre for Applied Biodiversity Science, Conservation International, Washington, DC, and the South African National Research Foundation for funding support. AGR was funded partly by NSF grant DEB0089801 and the National Centre for Ecological Analysis and Synthesis.

CITED REFERENCES

- Ackerly, D. D., S. A. Dudley, S. E. Sultan, J. Schmitt, J. S. Coleman, R. Linder, D. R. Sandquist, M. A. Geber, A. S. Evans, T. E. Dawson, and M. J. Lechowicz. 2000. The evolution of plant ecophysiological traits: recent advances and future directions. *BioScience* **50**:979-995.
- Anon. 1999. S-PLUS 2000 User's Guide. MathSoft, Seattle, WA.
- Austin, M. P., R. B. Cunningham, and P. M. Fleming. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* **55**:11-27.
- Austin, M. P., and T. M. Smith. 1989. A new model for the continuum concept. *Vegetatio* **83**:35-47.
- Bachelet, D., R. P. Neilson, J. M. Lenihan, and R. J. Drapek. 2001. Climate change effects on vegetation distribution and carbon budget in the United States. *Ecosystems* **4**:164-185.
- Bond, W. J. 1988. Proteas as "tumbleseeds": Wind dispersal through air and over soil. *South African Journal of Botany* **54**:455-460.
- Bond, W. J., and J. J. Midgley. 1988. Allometry and sexual differences in size. *The American Naturalist* **131**:909-910.
- Chapin III, S. F., E. S. Zavaleta, V. T. Eviner, R. L. Naylor, P. M. Vitousek, H. L. Reynolds, D. U. Hooper, S. Lavorel, O. E. Sala, S. E. Hobbie, M. C. Mack, and S. Diaz. 2000. Consequences of changing biodiversity. *Nature* **405**:234-242.
- Chessel, D., J. D. Lebreton, and R. Prodon. 1982. Mesures symétriques d'amplitude d'habitat et de diversité intra-échantillon dans un tableau espèces-relevés: cas d'un gradient simple. *Compte Rendu de l'Académie des Sciences de Paris D* **295**:83-88.
- Cornelissen, J. H. C. 1999. A triangular relationship between leaf size and seed size among woody species: allometry, ontogeny, ecology and taxonomy. *Oecologia* **118**:248-255.
- Díaz, S., and M. Cabido. 1997. Plant functional types and ecosystem function in relation to global change. *Journal of Vegetation Science* **8**:463-474.
- Díaz, S., M. Cabido, and F. Casanoves. 1998. Plant functional traits and environmental filters at a regional scale. *Journal of Vegetation Science* **9**:113-122.

- Doledec, S., D. Chessel, and C. Gimaret-Carpentier. 2000. Niche separation in community analysis: a new method. *Ecology* **81**:2914-2927.
- Doledec, S., D. Chessel, C. J. F. Ter Braak, and S. Champely. 1996. Matching species traits to environmental variables: a new three-table ordination method. *Environmental and Ecological Statistics* **3**:143-146.
- Fonseca, C. R., J. M. Overton, B. Collins, and M. Westoby. 2000. Shifts in trait-combinations along rainfall and phosphorus gradients. *Journal of Ecology* **88**:964-977.
- Givnish, T. J. 1987. Comparative studies of leaf form: assessing the relative roles of selective pressures and phylogenetic constraints. *New Phytologist* **106**:131-160.
- Goldblatt, P., and J. Manning. 2000. Cape Plants: a conspectus of the Cape Flora of South Africa. *Strelitzia* **9**:1-20.
- Green, R. H. 1971. A multivariate statistical analysis approach to the Hutchinson niche: bivalve molluscs of central Canada. *Ecology* **52**:543-556.
- Grime, J. P. 1974. Vegetation classification by reference to strategies. *Nature* **250**:26-31.
- Grime, J. P. 1979. Plant strategies and vegetation processes. John Wiley & Sons.
- Grime, J. P., K. Thompson, R. Hunt, J. G. Hodgson, J. H. C. Cornelissen, I. H. Rorison, G. A. F. Hendry, T. W. Ashenden, A. P. Askew, S. R. Band, R. E. Booth, C. C. Bossard, B. D. Campbell, J. E. L. Cooper, A. W. Davison, P. L. Gupta, W. Hall, D. W. Hand, M. A. Hannah, S. H. Hillier, D. J. Hodgkinson, A. Jalili, Z. Liu, J. M. L. Mackey, N. Matthews, M. A. Mowforth, A. M. Neal, R. J. Reader, K. Reiling, W. Ross-Fraser, R. E. Spencer, F. Sutton, D. E. Tasker, P. C. Thorpe, and J. Whitehouse. 1997. Integrated screening validates primary axes of specialisation in plants. *Oikos* **79**:259-281.
- Hill, M. O. 1974. Correspondence analysis: a neglected multivariate method. *Journal of the Royal Statistical Society. Series B, Statistical Methodological* **23**:340-354.
- Hodgson, J. G., P. J. Wilson, R. Hunt, J. P. Grime, and K. Thompson. 1999. Allocating C-S-R plant functional types: a soft approach to a hard problem. *Oikos* **85**:282-294.
- Huston, M. A. 1994. Biological diversity. Cambridge University Press, Cambridge.
- Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* **22**:145-159.
- Iverson, L. R., and A. Prasad. 1998. Predicting abundance for 80 tree species following climate change in the Eastern United States. *Ecological Monographs* **68**:465-485.
- Legendre, P., R. Galzin, and M. Harmelin-Vivien. 1997. Relating behavior to habitat: solutions to the fourth-corner problem. *Ecology* **78**:547-562.
- Malanson, G. P., W. E. Westman, and Y.-L. Yan. 1992. Realized versus fundamental niche functions in a model of chaparral response to climatic change. *Ecological Modelling* **64**:261-277.
- Midgley, G. F., L. Hannah, D. Millar, W. Thuiller, and A. Booth. 2003. Developing regional and species-level assessments of climate change impacts on biodiversity: A preliminary study in the Cape Floristic Region. *Biological Conservation* **112**:87-97.
- Niinemets, N. 2001. Global-scale climatic controls of leaf dry mass per area, density, and thickness in trees and shrubs. *Ecology* **82**:453-469.
- Parkhurst, D. F., and O. L. Loucks. 1972. Optimal leaf size in relation to environment. *Journal of Ecology* **60**:505-537.
- Prentice, I. C., W. Cramer, S. P. Harrison, R. Leemans, R. A. Monserud, and A. M. Solomon. 1992. A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography* **19**:117-134.
- Pulliam, H. R. 2000. On the relationship between niche and distribution. *Ecology Letters* **3**:349-361.
- Rebello, A. G. 1992. Protea Atlas Manual: instruction booklet to the Protea Atlas Project. National Botanical Institute, Claremont.
- Rebello, A. G. 2001. Proteas, a field guide to the Proteas of southern Africa. Fernwood Press, Vlaeberg, Cape Town.
- Reich, P. B., D. S. Ellsworth, M. B. Walters, J. M. Vose, C. Gresham, J. C. Vollin, and W. D. Bowman. 1999. Generality of leaf trait relationships: a test across six biomes. *Ecology* **80**:1955-1969.
- Rousseeuw, P. J., and A. M. Leroy. 1987. Robust Regression and Outlier Detection. Wiley, New York.
- Schulze, R. E. P. i., editor. , Junk, The Hague. 1973. Climatic indices and classifications in relation to the biogeography of southern Africa. Pages 19-54 in M. J. A. Werger, editor. *Biogeography and ecology of southern Africa*, Junk, The Hague.
- Smith, T. M., H. H. Shugart, and F. I. Woodward. 1997. Plant functional types. Cambridge Univ. Press, Cambridge.
- ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**:1167-1179.
- ter Braak, C. J. F. 1987. CANOCO-a FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis and redundancy analysis, Groep Landbouwwisckeunde, Wageningen, The Netherlands.
- Thioulouse, J., D. Chessel, S. Doledec, and J. M. Olivier. 1997. ADE4: a multivariate analysis and graphical display software. *Statistics and Computing* **7**:75-83.

- Thuiller, W. 2003. BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* **9**:000-000.
- Thuiller, W., J. Vaydera, J. Pino, S. Sabaté, S. Lavorel, and C. Gracia. 2003. Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology & Biogeography* **12**:313-325.
- Weiher, E., A. van der Werf, K. Thompson, M. Roderick, E. Garnier, and O. Eriksson. 1999. Challenging Theophrastus: A common core list of plant traits for functional ecology. *Journal of Vegetation Science* **10**:609-620.
- Westoby, M. 1998. A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil* **199**:213-227.
- Westoby, M., D. S. Falster, and A. T. Moles. 2002. PLANT ECOLOGICAL STRATEGIES: Some Leading Dimensions of Variation Between Species. *Annu. Rev. Ecol. Syst.* **33**:125-159.
- Williams, I. J. M. 1972. A revision of the genus *Leucadendron* (Proteaceae). *Contributions from the Bolus Herbarium*: 3.
- Woodward, F. I., and W. Cramer. 1996. Plant functional types and climate changes: Introduction. *Journal of Vegetation Science* **7**:306-308.
- Wright, I. J., P. B. Reich, and M. Westoby. 2001. Strategy shifts in leaf area physiology, structure and nutrient content between species of high- and low-rainfall and high- and low-nutrient habitats. *Functional Ecology* **15**:423-434.
- Wright, I. J., and M. Westoby. 1999. Differences in seedling growth behaviour among species: trait correlations across species, and trait shifts along nutrient compared to rainfall gradients. *Journal of Ecology* **87**:85-97.

Table 1: Percentages of explained inertia and cumulated explained inertia by axes following the multivariate analysis.

	Explained inertia			Cumulated explained inertia		
	OMI	CCA	RDA	OMI	CCA	RDA
Axis 1	46%	26%	34%	46%	34%	26%
Axis 2	27%	24%	28%	73%	62%	51%
Axis 3	21%	15%	15%	94%	77%	66%

Table 2: Canonical correlation between bio-climatic variables and the first three axes for each multivariate analysis.

	OMI			CCA			RDA		
	Axis 1	Axis 2	Axis 3	Axis 1	Axis 2	Axis 3	Axis 1	Axis 2	Axis 3
Potential evapotranspiration	-0.946	-0.051	-0.088	-0.722	0.251	0.469	0.770	0.253	0.379
Soil moisture in summer	0.244	0.501	-0.244	0.848	0.320	-0.151	-0.823	0.330	-0.349
Soil moisture in winter	0.393	-0.834	0.321	-0.180	-0.678	-0.506	0.055	-0.865	-0.243
Mean annual precipitation	0.905	-0.132	0.201	0.393	-0.098	-0.568	-0.484	-0.268	-0.596
Summer precipitation	0.728	0.490	-0.278	0.757	0.396	-0.092	-0.741	0.410	-0.317
Winter precipitation	0.694	-0.565	0.341	0.048	-0.450	-0.598	-0.171	-0.651	-0.466
Mean annual temperature	-0.452	-0.233	-0.915	0.090	-0.285	0.828	-0.149	0.119	0.838
Max T° of the warmest month	-0.759	0.133	-0.153	-0.481	0.368	0.426	0.498	0.381	0.271
Max T° of the coldest month	0.077	-0.538	-0.761	0.294	-0.711	0.524	-0.420	-0.345	0.741
Heat units	-0.636	-0.205	-0.695	-0.228	-0.084	0.738	0.182	0.178	0.668

FIGURE CAPTIONS

Figure 1: Ordination diagrams on the first two axes of OMI analysis. A) Representation of the explicative variables in the OMI plan 1-2; B) representation of the explicative variables in the OMI plan 1-3.

Figure 2: Species niche position and niche breadth on the OMI axes for *Leucadendron* in the Cape Floristic Region. A) axis 1: aridity gradient; B) axis 2: seasonality of precipitation; C) axis 3: temperature. The horizontal bars correspond to standard deviations and are interpreted as niche breadths. Small vertical bars at the bottom of each pane correspond the position of sites along the axes.

Figure 3: Geographical variation in the first two OMI axes scores (axis 1: aridity gradient; axis 2: seasonality of precipitation) amongst the samples in Cape Floristic region. Both OMI axes were overlaid, showing regional differences in the overall variation between the OMI axes. Black grid cells show low scores for both axes, white show high scores for both. Increasing blue color represents increasing scores of aridity gradient and increasing intensities of green represents decreasing scores of seasonality of precipitation. Shoot designs and allometrically related aspects associated with moisture availability are indicated alongside these axes, with larger coned, larger leaved species associated with more humid climates.

Figure 4: Difference between *Leucadendron* species following the dispersion mode on the three OMI axes. A) axis 1: aridity gradient; B) axis 2: seasonality of precipitation; C) axis 3: temperature. The only significant differences are for seasonality of precipitation between species dispersed by wind and ants ($P=0.008$). Lowercase letters are placed above the bars that were compared using Wilcoxon tests. Bars with the same letters are not significantly different ($P>0.05$)

Figure 5: Linear relation between the plant leaf area and the first OMI axis: aridity gradient for *Leucadendron* in the Cape Floristic Region.

Figure 6: Relation between plant height (A) and leaf area (B) and plant dispersion mode for *Leucadendron* in the Cape Floristic Region.

Figure 7: Difference between the onsets of flowering (A-C) and ends of flowering (D-F) for *Leucadendron* species in the Cape Floristic Region on the first three OMI axes. A & D) OMI axis 1: aridity gradient; B & E) OMI axis 2: seasonality of precipitation gradient; C & F) OMI axis 3: temperature gradient.

Figure 2-A

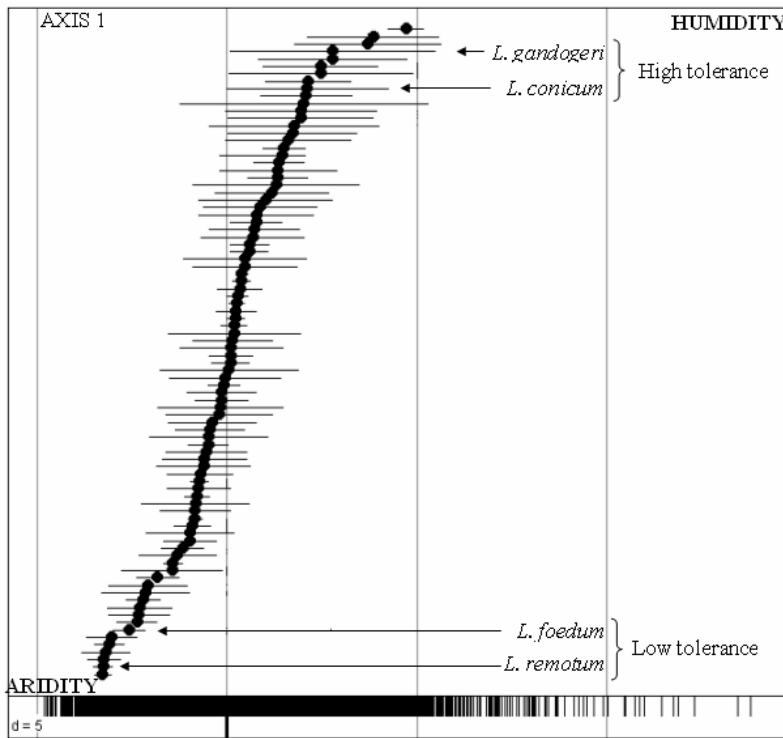


Figure 2-B

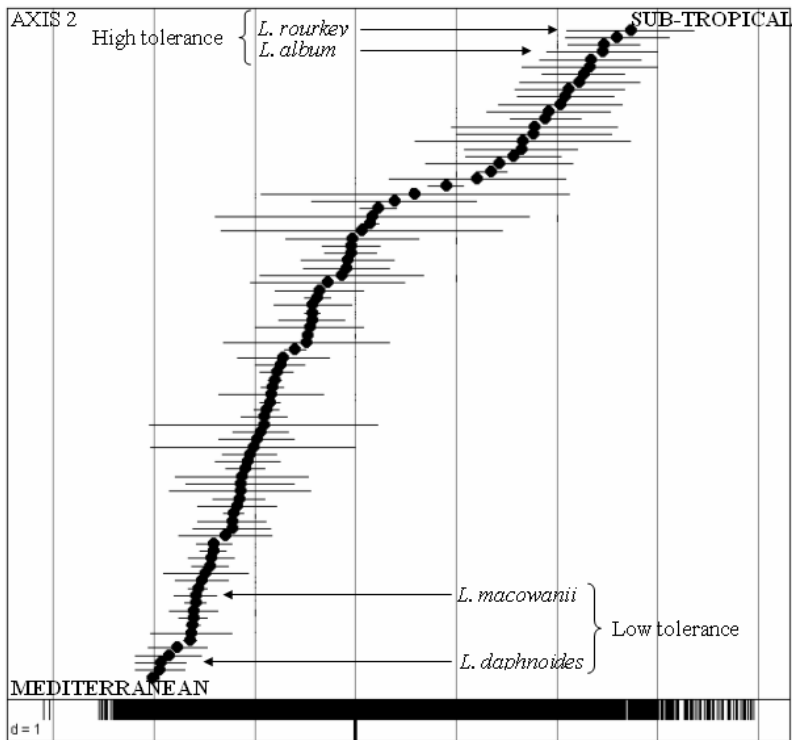


Figure 2-C

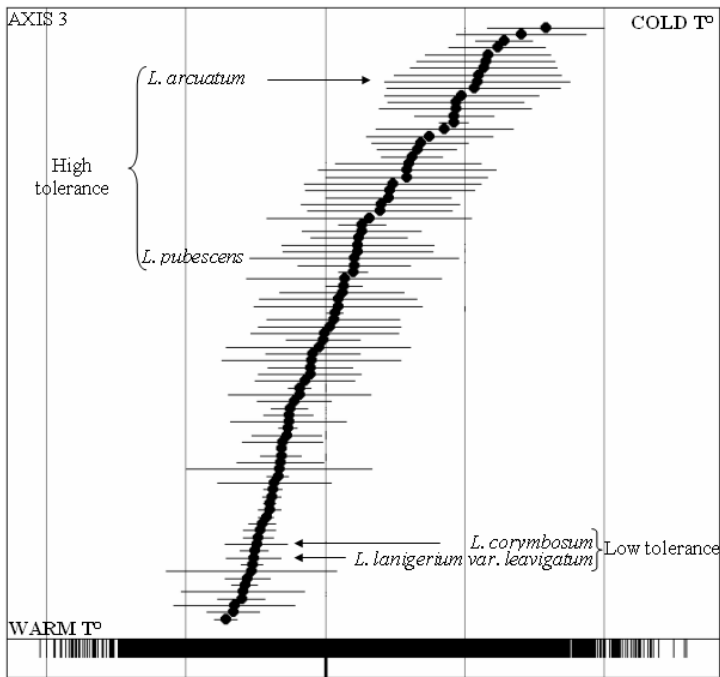


Figure 1-A

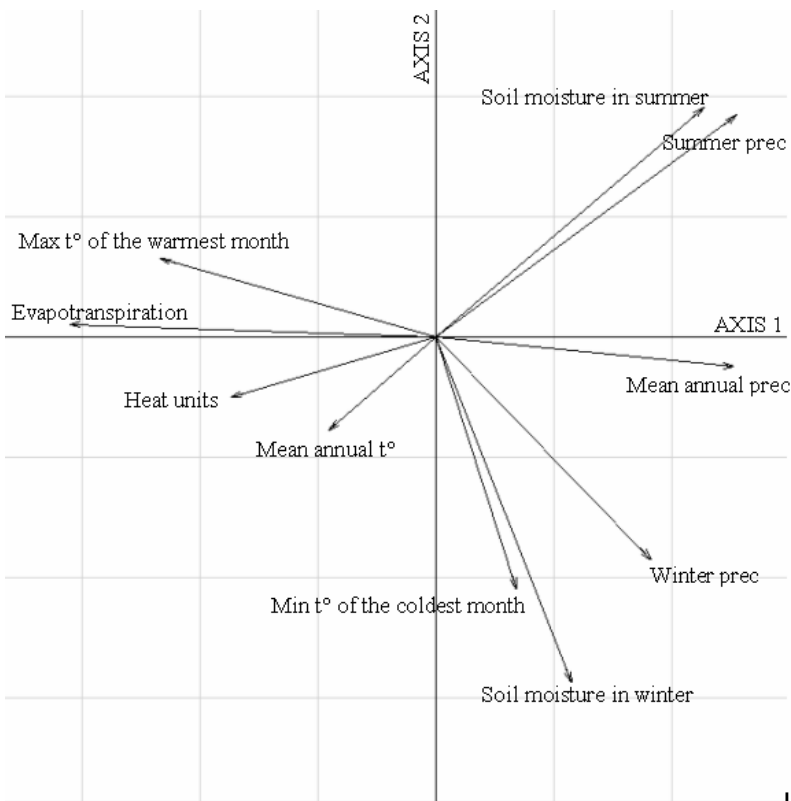


Figure 1-B

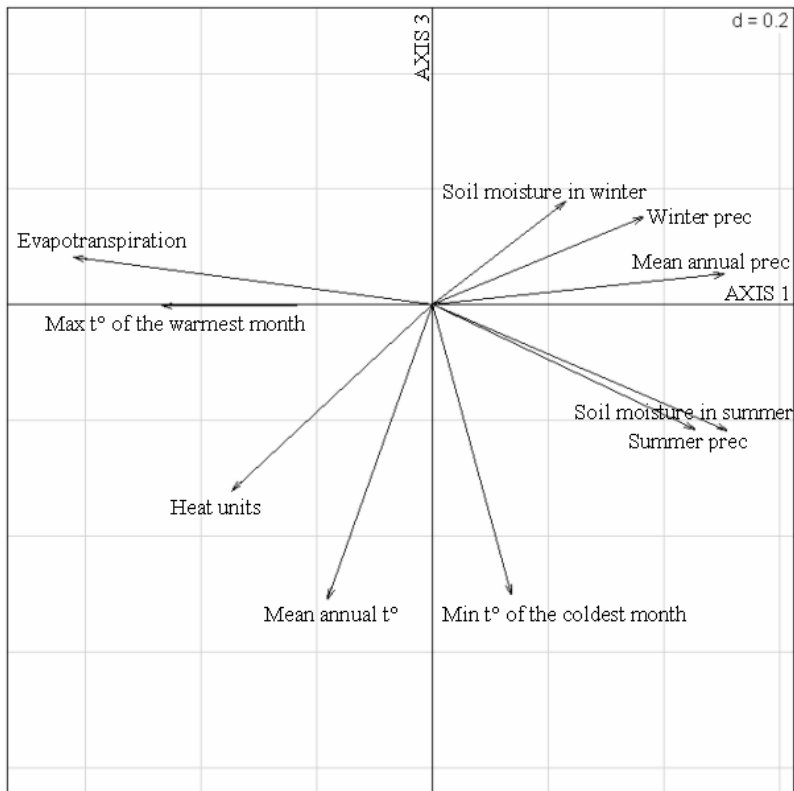


Figure 3

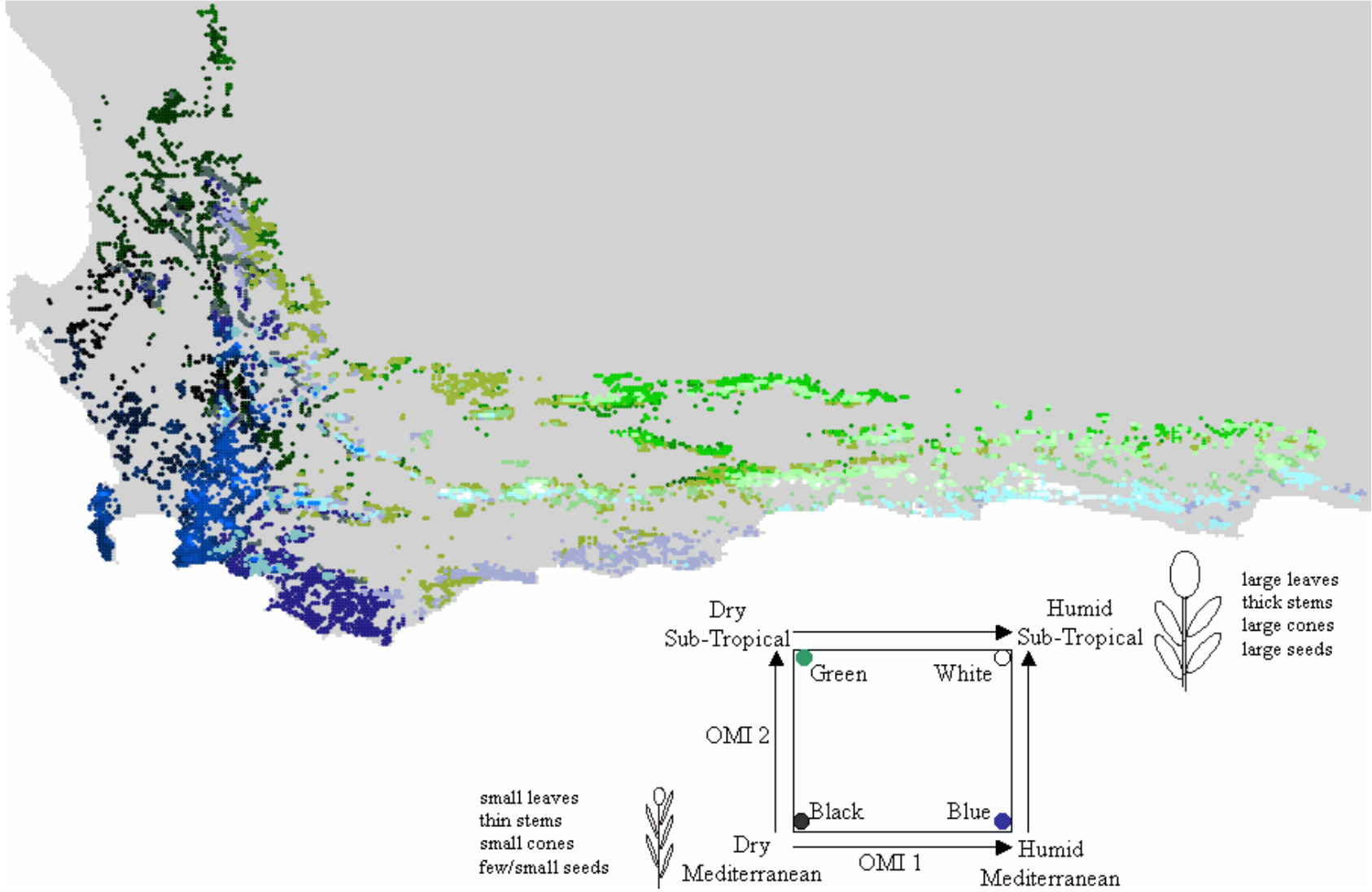


Figure 4

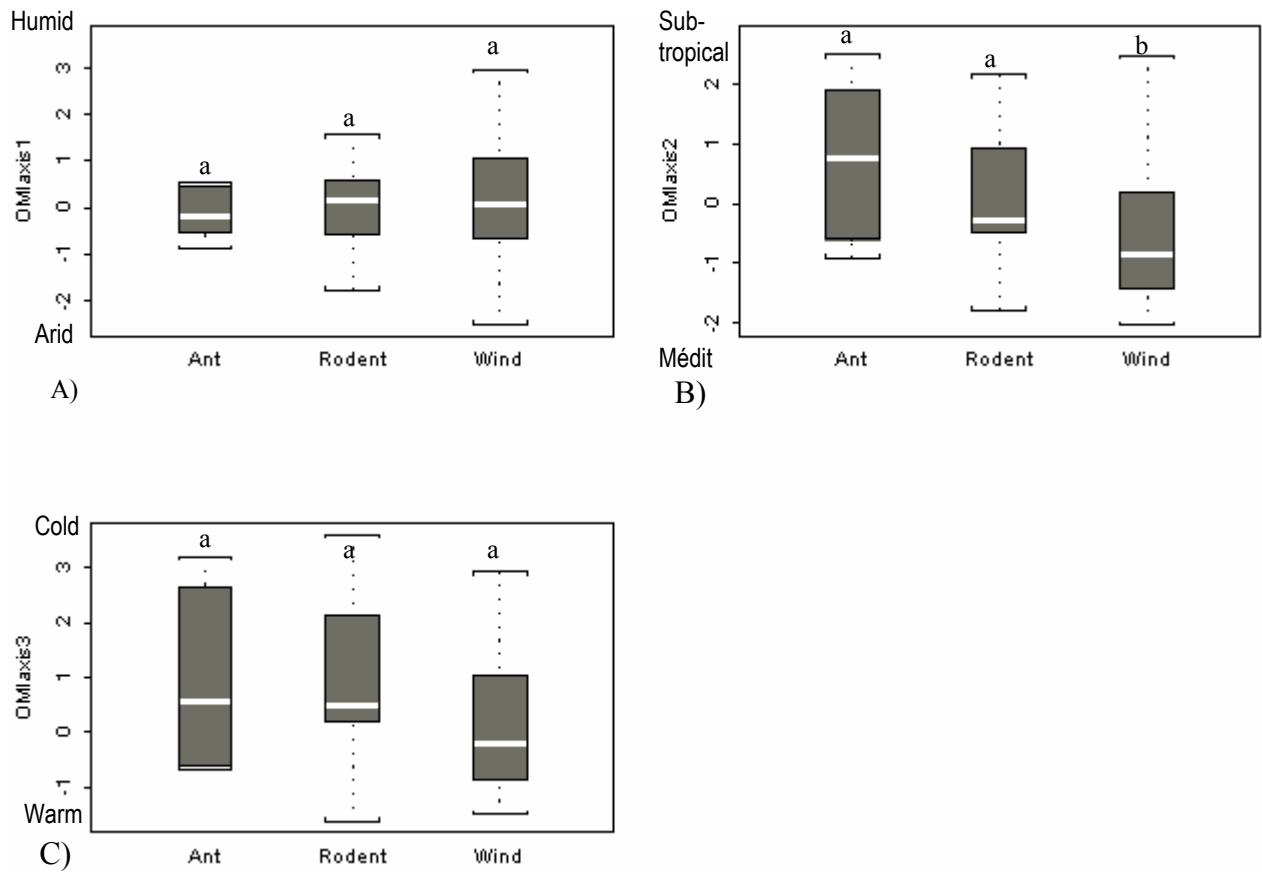


Figure 5

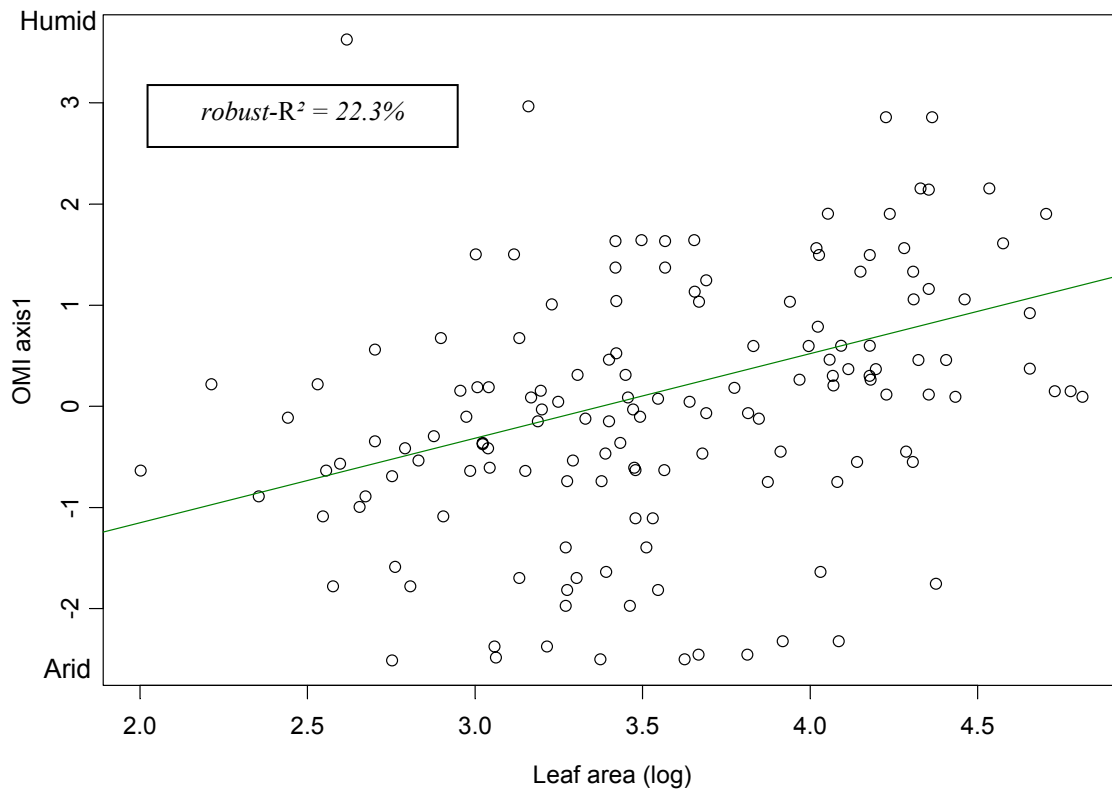
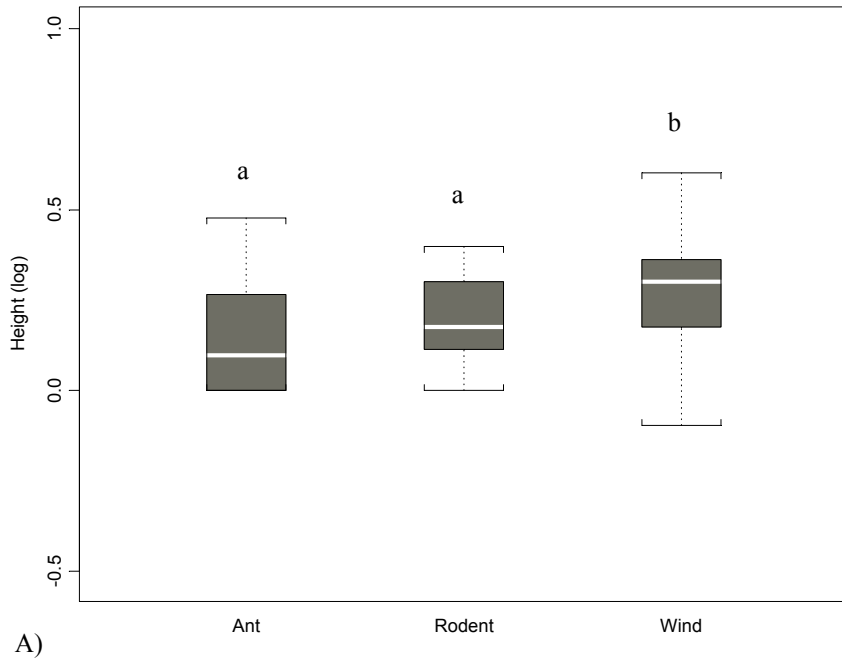
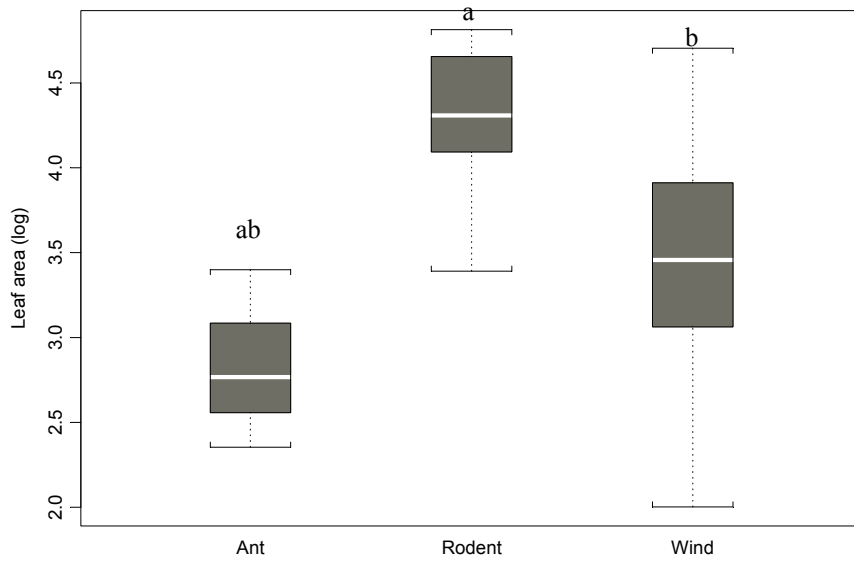


Figure 6



A)



B)

Figure 7

