UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Unfolding nature *in silico* – Overcoming practical and methodological limitations of species distribution models

Applications to conservation biology and climate change impact assessment

Thèse de doctorat ès sciences de la vie (PhD)

Présentée à la

Faculté de Biologie et Médecine de l'Université de Lausanne

Par

Robin Engler

Diplômé en Biologie (Université de Lausanne)

Jury de Thèse:

Prof. Yves Poirier – Président
Prof. Antoine Guisan – Directeur de thèse
Prof. Thomas C. Edwards – Expert
Dr. Stefan Dullinger – Expert

Lausanne 2009

UNIL | Université de Lausanne
Faculté de biologie
et de médecine
Département d'Écologie et d'Évolution

Unfolding nature *in silico* – Overcoming practical and methodological limitations of species distribution models

Applications to conservation biology and climate change impact assessment

Thèse de doctorat ès sciences de la vie (PhD)

Présentée à la

Faculté de Biologie et Médecine de l'Université de Lausanne

Par

Robin Engler

Diplômé en Biologie (Université de Lausanne)

Jury de Thèse:

Prof. Yves Poirier – Président
Prof. Antoine Guisan – Directeur de thèse
Prof. Thomas C. Edwards – Expert
Dr. Stefan Dullinger – Expert

Lausanne 2009

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président	Monsieur	Prof.	Yves	**Poirier**
Directeur de thèse	Monsieur	Prof.	Antoine	**Guisan**
Experts	Monsieur	Prof.	Thomas C.	**Edwards**
	Monsieur	Dr	Stefan	**Dullinger**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Robin Engler

biologiste diplômé de l'Université de Lausanne

intitulée

Unfolding nature *in silico*
Overcoming practical and metodological limitations
of species distribution models

Lausanne, le 6 novembre 2009

pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Yves Poirier

Abstract

The existence of a causal relationship between the spatial distribution of living organisms and their environment, in particular climate, has been long recognized and is the central principle of biogeography. In turn, this recognition has led scientists to the idea of using the climatic, topographic, edaphic and biotic characteristics of the environment to predict its potential suitability for a given species or biological community.

In this thesis, my objective is to contribute to the development of methodological improvements in the field of species distribution modeling. More precisely, the objectives are to propose solutions to overcome limitations of species distribution models when applied to conservation biology issues, or when used as an assessment tool of the potential impacts of global change.

The first objective of my thesis is to contribute to evidence the potential of species distribution models for conservation-related applications. I present a methodology to generate pseudo-absences in order to overcome the frequent lack of reliable absence data. I also demonstrate, both theoretically (simulation-based) and practically (field-based), how species distribution models can be successfully used to model and sample rare species. Overall, the results of this first part of the thesis demonstrate the strong potential of species distribution models as a tool for practical applications in conservation biology.

The second objective this thesis is to contribute to improve projections of potential climate change impacts on species distributions, and in particular for mountain flora. I develop and a dynamic model, MIGCLIM, that allows the implementation of dispersal limitations into classic species distribution models and present an application of this model to two virtual species. Given that accounting for dispersal limitations requires information on seed dispersal distances, a general methodology to classify species into broad dispersal types is also developed. Finally, the MIGCLIM model is applied to a large number of species in a study area of the western Swiss Alps. Overall, the results indicate that while dispersal limitations can have an important impact on the outcome of future projections of species distributions under climate change scenarios, estimating species threat levels (e.g. species extinction rates) for a mountainous areas of limited size (i.e. regional scale) can also be successfully achieved when considering dispersal as unlimited (i.e. ignoring dispersal limitations, which is easier from a practical point of view).

Finally, I present the largest fine scale assessment of potential climate change impacts on mountain vegetation that has been carried-out to date. This assessment involves vegetation from 12 study areas distributed across all major western and central European mountain ranges. The results highlight that some mountain ranges (the Pyrenees and the Austrian Alps) are expected to be more affected by climate change than others (Norway and the Scottish Highlands). The results I obtain in this study also indicate that the threat levels projected by fine scale models are less severe than those derived from coarse scale models. This result suggests that some species could persist in small refugias that are not detected by coarse scale models.

Résumé

L'existence d'une relation causale entre la répartition des espèces animales et végétales et leur environnement, en particulier le climat, a été mis en évidence depuis longtemps et est un des principes centraux en biogéographie. Ce lien a naturellement conduit à l'idée d'utiliser les caractéristiques climatiques, topographiques, édaphiques et biotiques de l'environnement afin d'en prédire la qualité pour une espèce ou une communauté.

Dans ce travail de thèse, mon objectif est de contribuer au développement d'améliorations méthodologiques dans le domaine de la modélisation de la distribution d'espèces dans le paysage. Plus précisément, les objectifs sont de proposer des solutions afin de surmonter certaines limitations des modèles de distribution d'espèces dans des applications pratiques de biologie de la conservation ou dans leur utilisation pour évaluer l'impact potentiel des changements climatiques sur l'environnement.

Le premier objectif majeur de mon travail est de contribuer à démontrer le potentiel des modèles de distribution d'espèces pour des applications pratiques en biologie de la conservation. Je propose une méthode pour générer des pseudo-absences qui permet de surmonter le problème récurent du manque de données d'absences fiables. Je démontre aussi, de manière théorique (par simulation) et pratique (par échantillonnage de terrain), comment les modèles de distribution d'espèces peuvent être utilisés pour modéliser et améliorer l'échantillonnage des espèces rares. Ces résultats démontrent le potentiel des modèles de distribution d'espèces comme outils pour des applications de biologie de la conservation.

Le deuxième objectif majeur de ce travail est de contribuer à améliorer les projections d'impacts potentiels des changements climatiques sur la flore, en particulier dans les zones de montagnes. Je développe un modèle dynamique de distribution appelé MIGCLIM qui permet de tenir compte des limitations de dispersion dans les projections futures de distribution potentielle d'espèces, et teste son application sur deux espèces virtuelles. Vu que le fait de prendre en compte les limitations dues à la dispersion demande des données supplémentaires importantes (p.ex. la distance de dispersion des graines), ce travail propose aussi une méthode de classification simplifiée des espèces végétales dans de grands "types de disperseurs", ce qui permet ainsi de d'obtenir de bonnes approximations de distances de dispersions pour un grand nombre d'espèces. Finalement, j'applique aussi le modèle MIGCLIM à un grand nombre d'espèces de plantes dans une zone d'études des pré-Alpes vaudoises. Les résultats montrent que les limitations de dispersion peuvent avoir un impact considérable sur la distribution potentielle d'espèces prédites sous des scénarios de changements climatiques. Cependant, quand les modèles sont utilisés pour évaluer les taux d'extinction d'espèces dans des zones de montages de taille limitée (évaluation régionale), il est aussi possible d'obtenir de bonnes approximations en considérant la dispersion des espèces comme illimitée, ce qui est nettement plus simple d'un point de vue pratique.

Pour terminer je présente la plus grande évaluation à fine échelle d'impact potentiel des changements climatiques sur la flore des montagnes conduite à ce jour. Cette évaluation englobe 12 zones d'études réparties sur toutes les chaines de montages principales d'Europe occidentale et centrale. Les résultats montrent que certaines chaines de montagnes (les Pyrénées et les Alpes Autrichiennes) sont projetées comme plus sensibles aux changements climatiques que d'autres (les Alpes Scandinaves et les Highlands d'Ecosse). Les résultats obtenus montrent aussi que les modèles à échelle fine projettent des impacts de changement climatiques (p. ex. taux d'extinction d'espèces) moins sévères que les modèles à échelle large. Cela laisse supposer que les modèles a échelle fine sont capables de modéliser des micro-niches climatiques non-détectées par les modèles à échelle large.

Acknowledgements

First and foremost I would like to express all my gratitude to Antoine Guisan for giving me the opportunity, and pushing me, to carry-out this thesis under his supervision. The combination of freedom, knowledge and guidance that he offered was a key to the successful completion of this thesis and led us to what I believe has been a very fruitful collaboration.

I am also greatly thankful to Stefan Dullinger and Thomas Edwards who have kindly agreed to expertise this thesis and have also provided highly valuable comments on a number of manuscripts included in this thesis at an earlier stage.

Among the people without whom this thesis would simply not have been possible, I am particularly indebted towards Christophe Randin and Gwenaëlle Le Lay. They have provided me with key data and collaborations that have allowed me to achieve the present work. They are not only highly competent and extraordinarily motivated scientists, but most importantly also great friends.

I would also like to thank Pascal Vittoz, Olivier Broennimann, Peter Pearman, Nicklaus Zimmermann, Wilfried Thuiller, Miguel Araújo, Nigel Yoccoz and Martin Beniston with whom I have collaborated on different aspects of this thesis and whose knowledge and expertise have been of invaluable help.

I also wish to thank Thomas Czáka for his assistance with, and passion for, geographical information systems that has been so helpful, especially at the earlier stages of my thesis.

My thanks also go to Erica Frank and Florian Dessimoz who carried-out their master thesis in the spatial ecology group and with whom I collaborated.

I would also like to thank all the members of the spatial ecology team that I have not acknowledged at this stage. In alphabetic order of surname: Carmen Cianfrani, Anne Dubuis, Luigi Maiorano, Loïc Pellissier, Blaise Petitpierre, Dorothea Pio, Patricio Pliscoff, Julien Pottier and Charlotte N'diribe. The daily lab-life with them was a pleasure and I will certainly miss this working atmosphere.

I also wish to acknowledge the department of ecology of evolution of the University of Lausanne who provided a stimulating yet relaxed working environment.

On a less scientific note, I would also like to express my gratitude to my office mates Dorothea Pio, Gwenaëlle Le Lay and Julien Pottier. Sharing the office with them was a lot of fun and certainly helped in those moments when papers were coming back from review with not exactly the praise that I was hoping for.

Last but not least, I would like to thank all my friends, and in particular Luca Rechsteiner, my girlfriend Jihyun Lee and my parents and brothers for their support as well as for forcing me to leave my computer screen, sometimes.

Table of Content

Introduction

Species distribution models: unfolding nature in silico

Knowledge about the spatial distribution of animal and plant species has certainly been central to humans since the dawn of time. Early *Homo sapiens* were probably already interested to know the locations of edible plants or where the likelihood of coming across the ill-fated game that would become their next lunch was maximal. Possibly, they might already have been able to relate the presence or absence of certain species with particular environmental conditions.

This existence of a causal relationship between the spatial distribution of plants and animals and their environment, in particular climate, has been long recognized and is the central principle in the field of biogeography (Lischke et al. 1998a, Guisan and Zimmermann 2000, Pearson and Dawson 2003, and references cited therein). In turn, this recognition has led scientist to the idea of using the climatic, topographic, edaphic and biotic characteristics of the environment to predict its potential suitability for a given species or biological community. To project the potential distribution of species and more particularly that of plants, different categories of models exist. Most can however be categorized as either mechanistic or empirical (Lischke et al. 1998a).

The first category, mechanistic models (also sometimes called "process-based" or "physiological"), attempt to model ecophysiological processes in as much detail as possible. In this category, we find for instance the so called "gap models" (Botkin et al. 1972) that are used to model forest growth and succession (e.g. Sykes at al. 1996, Lischke et al. 1998b, Rickbush et al. 2007) as well as "dynamic global vegetation models" (see Woodward and Lomas 2004 for a review) that are mostly employed to model plant functional types (i.e., broad vegetation units) at a continental scale (e.g. Prentice et al. 1992, Woodward et al. 1995, Cramer et al. 2001).

The second category, empirical models (also sometimes called "correlative" or "statistical"), are seeking to establish relationships between observed species occurrence and environmental factors that are of a statistical nature and hence do not necessarily imply a causal relationship. Thus, unlike mechanistic models, empirical models do not rely on underlying ecological functions and mechanisms (Franklin 1995, Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Hirzel and Le Lay 2008). In the literature, these correlative models are given a large number of different names, such as species distribution models, predictive habitat distribution models, habitat suitability models, bioclimate envelope models or resource selection functions to cite a few. However, in recent literature, the appellation "Species Distribution Models" (abbreviated SDM) seems to prevail. Although it can be argued whether this choice of terminology is really optimal, since, at least in their basic implementation, SDMs do model something closer to habitat suitability than to actual distribution, I will nevertheless employ this terminology to refer to these models throughout this introduction.

The concept of ecological niche

At the core of species distribution models (SDMs) is the concept of "ecological niche", which is the theoretical framework to the quantification of the relationship between species and their environment. Given the important body of literature that already discusses and reviews the concept of ecological niche in relationships with species distribution models (Pulliam 2000, Chase and Leibold 2003, Araújo and Guisan 2006, Pearman et al. 2008a, Hirzel and Le Lay

2008), I will here simply provide an overview focusing of the most relevant aspects of ecological niche theory in relation to SDMs.

The concept of ecological niche as used in SDMs was formalized by Hutchinson (1957) as an "n-dimensional hypervolume" in the space of the environmental variables that permit indefinite survival of a species. In other words, the ecological niche of a species, as defined by Hutchinson, is the ensemble of environmental conditions under which a species can maintain a positive growth rate. Hutchinson refined this concept by further introducing the distinction between "fundamental" and "realized" niche. While the former only accounts for abiotic factors such as climate, topography or soil proprieties, the latter also accounts for biotic interactions such as competition or facilitation. A classical illustration of fundamental vs. realized niche is given by the distribution limit of mountain plants, which are generally limited by climatic conditions at the higher end of their range, and by competition at the lower end (Brown et al. 1996). For such species, the lower elevation parts of mountains are thus part of their fundamental niche (i.e. they could occupy these habitats in the absence of competition) but not of their realized niche (i.e. they are absent because outcompeted from these habitats).

Since they are calibrated from field observations of species that include the effects of biotic interactions rather than from physiological measures obtained under controlled experimental conditions, SDMs mostly capture something that is close to the realized niche of a species. The terms "mostly" and "close" are here important because different SDM techniques are thought to have different positions along the fundamental-realized niche gradient (Jimenez-Valverde et al. 2008).

While most modern SDMs are still based on the Hutchinsonian concept of the niche, some authors (e.g. Pulliam 2000) have highlighted that it takes more than environmental habitat suitability and competition to determine the presence or absence of a species from a given habitat. Dispersal limitation, local extinctions or re-colonization are frequent in nature and can result in a species being absent from suitable habitat and present in unsuitable habitat (Pulliam 2000, and references cited therein).

This means that meta-population and source-sink dynamics such as dispersal and population extinctions should also be incorporated into the concept of niche and are at least as important as competition in shaping the ecological niches of species. For instance, Pulliam (2000) cites examples where dispersal limitations result in species being absent of some suitable habitat, and, conversely, where populations are maintained despite having a negative growth rate. Pulliam also shows, through modeling, how dispersal can maintain populations in unsuitable areas, i.e. areas where a species' growth rate is < 1. If dispersal is high, simulations show that a significant proportion of individuals (up to ~30%) can live outside of their fundamental niche. However, as to date, most SDMs applications ignore these meta-population and source-sink dynamics.

SDMs assume that the best representation of a species' environmental requirements is its current observed distribution (Pearson and Dawson 2003). This implies that, if the full potential distribution of a species is to be modeled, a species needs to be in equilibrium with its environment. This "equilibrium" or "pseudo-equilibrium" between observed distribution and environment is therefore a fundamental assumption of SDMs (Guisan and Zimmermann 2000), and, for this reason, SDMs are also sometimes referred to as "equilibrium models". A typical example of non-equilibrium situation is that of an invasive species newly introduced to a territory. By definition, such a species is in a colonization phase and does thus not occupy all of its suitable habitats. Another example is that of many European trees that are thought to fill only a limited part of their potentially suitable habitat because they have not completed yet their post-glacial re-colonization from their ice-age refugia (Svenning and Skov 2004). Again,

the species are not in equilibrium with their environment because they area missing from some of their potentially suitable habitats.

Another important assumption of SDMs is "niche conservatism" (Peterson et al. 1999, see Pearman et al. 2008 for a review). However, this assumption only becomes of concern when a model is transferred in space and/or time, and is therefore discussed in a next section of this introduction.

Applications of species distribution models

A search on the ISI Web of Science (www.isiknowledge.com; 16.07.2009) for the key words "'species distribution' model", "'habitat suitability' model", "'ecological niche' model" and "'habitat suitability' map" returned close to 2000 publications in the field of ecology over the past three decades (Figure 1a). The increase in publications is particularly steep over last 10 years or so, and goes beyond what can be explained by the general increase in scientific publications and in particular in the field of ecology (Figure 1b).

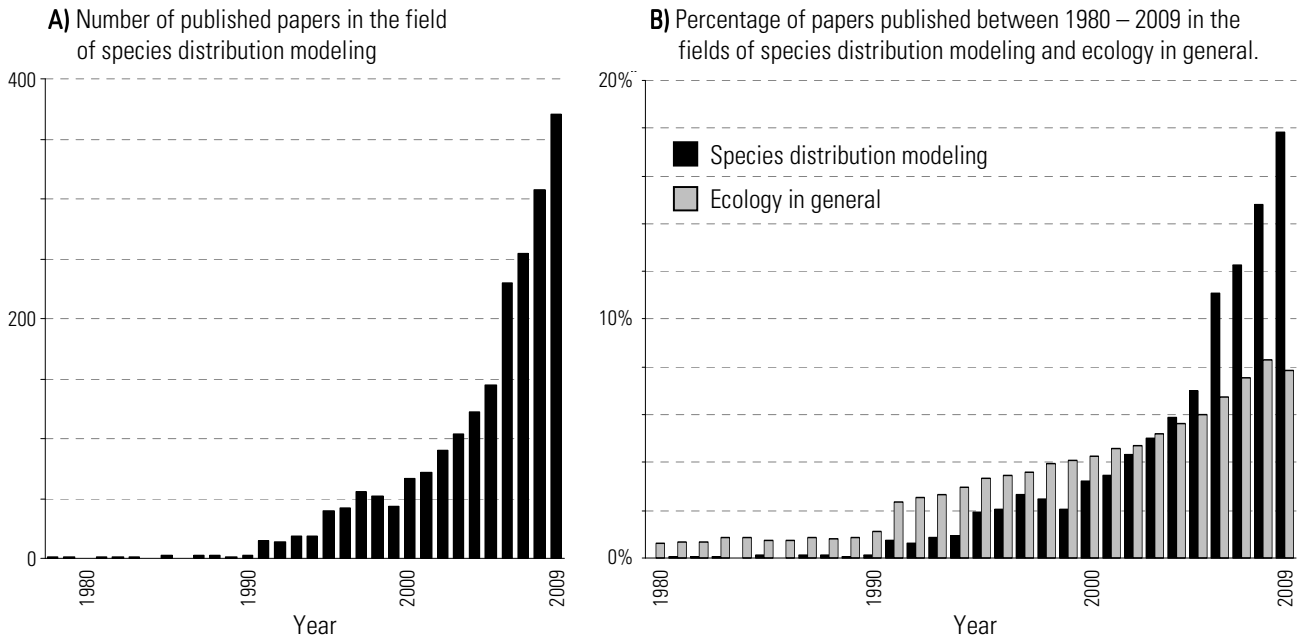


Figure 1. A) Number of scientific publications published per year between 1978 and 2009 in the field of species distribution modeling. B) Comparison of the proportion (given as a percentage) of scientific publications published per year between 1980 and 2009 between the fields of species distribution modeling and ecology in general. Values for 2009 are projections based on the first half of the year.

Part of the explanation for this trend is certainly to be found in the increase of computing power, sophistication of software (in particular statistical and geographical information systems software), increased availability of geo-referenced species data (e.g. Global Biodiversity Information Facility, www.gbif.org) and environmental data (e.g. worldclim data, Hijmans et al. 2005), as well as the development of ever more user-friendly interfaces for

model calibration and projection (e.g. BIOMOD, Thuiller et al. 2009; MAXENT, Philipps et al. 2005; or BIOMAPPER, Hirzel et al. 2002). But this trend also reflects genuine increase in interest from the scientific community for these tools that can be used to tackle both practical and theoretical issues in ecology.

Given the sheer number of publications, it is not surprising that SDMs have been applied to almost any form of living organisms. To cite only a few, models have been developed for submerged vegetation (Lehmann 1998), lichens (Edwards et al. 2005), fungi (Wollan et al. 2008), ferns (Zaniewski et al. 2002) and a large number of higher plants (e.g. Austin 1996, Dullinger et al. 2004).

Animal species are also well represented, with models calibrated for corals (Tittensor et al. 2009), mollusks (Mueller et al. 2009), crustaceans (Galparsoro et al. 2009, Murase et al. 2009), insects such as ants (Maggini et al. 2002) or butterflies (Luoto et al. 2006, Lutolf et al. 2006, Peterson 2007), spiders (Richardson et al. 2006, Acosta 2008) and mites (Estrada-Pena and Venzal 2007), fish (Lassale and Rochard 2009, Maxwell et al. 2009), amphibians (Pellet et al. 2004, Pineda and Lobo 2009, Brown et al. 2009), reptiles (Guisan and Hofer 2003), birds (Austin et al. 1996, Zarnetske et al. 2007, Lawler and Edwards 2002) and a multitude of mammals ranging from bats (Jaberg and Guisan 2001) to the giant panda (Qi et al. 2009).

Furthermore, SDMs have not only been used to project the distribution of individual species, but also that of entire communities (e.g. forest communities, Brzeziecki et al. 1995, Leathwick et al. 2001, Maggini et al. 2006), species richness (Iverson and Prasad 1998a, Guisan and Theurillat 2000a, Luoto et al. 2002, Mac Nally 2003, Thuiller et al. 2006) or endemicity (Raes et al. 2009).

Finally, the domains of applications for SDMs are equally numerous and include (see also Peterson 2006a for a review): Theoretical issues such as testing the conservatism of ecological niche in space and time (Peterson et al. 1999, Pearman et al. 2008a), understanding the ecological requirements of species (Austin and Meyers 1996) or their taxonomic relationships (Raxworthy et al. 2007), conservation biology applications such as finding new populations of rare (Edwards et al. 2005) or unknown species (Raxworthy et al. 2003), identifying suitable sites for translocation or reintroduction or designing nature reserves (Mladenoff et al. 1995), predict potential risk of species invasions (Peterson 2003, Broennimann et al. 2007), distribution of disease-transmitting species that jeopardize human health (Peterson 2006b, Peterson et al. 2006) and the forecast of potential impacts of climate change (Bakkenes et al. 2002, Thomas et al. 2004, Thuiller et al. 2005).

Species distribution models in practice

The process of species distribution model building has already been covered in detail in the existing literature (Franklin 1995, Lischke et al. 1998a, Guisan and Zimmermann 2000, Ferrier et al. 2002a, Ferrier et al. 2002b, Guisan and Thuiller 2005). Therefore, my only ambition here is to provide a concise summary of the main steps involved in the practical implementation of these models for those unfamiliar with this field. Leaving aside the conceptual aspects of model calibration (see Guisan and Zimmermann 2000), the practical implementation of species distribution models (SDMs) can be seen as composed of three main steps: (i) collection and preparation of the data to be used in model calibration (ii) modeling of the ecological niche, i.e. model calibration, and (iii) projection and evaluation of the model.

Data collection and preparation

Two types of data are needed for calibrating a SDM. Response variable data (i.e. species presence, absence or abundance) and explanatory variable data (i.e. environmental conditions at the response variable's locations).

The response variable is the quantity to be modeled, usually the presence and absence or the abundance of a species. But higher ensembles such as communities or assemblages can also be modeled (Ferrier et al. 2002b).

The explanatory variables are the environmental variables that are to be put in relation with the response variable. They are often referred to as "predictors". In the large majority of cases, studies have been using climate and topography variables because these are the most readily available. But other variables, including landuse (e.g. Randin et al. *in press*), NDVI (e.g. Chapter 1.3 of this thesis) or edaphic factors (e.g. Leathwick 1995, Iverson and Prasad 1998b), have also been employed successfully. The choice of variables is also driven by the spatial accuracy at which the modeling is to be carried-out. Fine scale models (typically ~20 – 100 m pixel size) will often benefit from including high resolution data (e.g. high resolution topography; Lassueur et al. 2006), while large scale studies (typically ~1 km – 50 km pixel size) will mostly be based solely on climatic variables.

Finally, environmental variables are also classified as being either proximal (direct), if there is a causal relationship between them and the species' occurrence, or distal (indirect) if this relationship is indirect (Austin and Meyers 1996). The most trivial examples to illustrate these two categories are temperature, which is a proximal variable because it has a causal effect on a species' fitness, and elevation, which is a distal variable as it only has an indirect effect a species' fitness (e.g. higher elevations are associated to lower temperatures).

Indirect variables are often an attractive option when they can be obtained with lower effort and more accuracy. For instance, elevation, in the form of a digital elevation model, is nowadays available worldwide with an accuracy of 30 meters (www.gdem.aster.ersdac.or.jp) and often much more from national topographic offices. From a biological point of view however, direct environmental variables should be favored because they are ecologically more meaningful and make models easier to interpret. Furthermore, indirect variables must be completely disregarded if the model is to be transposed to another geographical location (Austin and Meyers 1996).

Modeling of the ecological niche

This is the step where the actual model is fitted to the data, i.e. the ecological niche of the species is quantified in the space of the environmental variables. A non-exhaustive list of modeling techniques is given in table 1 here below (see also Guisan and Thuiller 2005).

Table 1. List of some of the most widely used techniques for species distribution modeling.
***** indicates methods for which only presence data are required.

Modeling methodology	Reference
Environmental Envelope (often referred to as BIOCLIM) *	Box 1981, Busby 1991
Classification Tree Analysis (CTA)	Breiman et al. 1984
Generalized Linear Models (GLM)	McCullagh and Nelder 1989; Guisan et al. 2002 for applications in ecology
Generalized Additive Models (GAM)	Hastie and Tibshirani 1986
Multivariate Adaptive Regression Splines (MARS)	Friedman 1991
Mixture Discriminant Analysis (MDA)	Hastie et al. 1994
Artificial Neural Networks (ANN	Ripley 1996
Generalized Boosted Models (GBM)	Ridgeway 1999
Random Forest (RF)	Breiman 2001
Genetic algorithms (GARP)	Peterson 2001
Ecological Nice Factor Analysis (ENFA) *	Hirzel et al. 2002
Maximum Entropy (MAXENT) *	Philipps et al. 2006

While quite a few studies have been dedicated to compare the relative performances of different modeling methods (e.g. Hirzel et al. 2001, Segurado and Araújo 2004, Elith et al. 2006, Guisan et al. 2007, Meynard and Quinn 2007, Tsoar et al. 2007, Elith and Graham 2009) indicating that some might be more powerful than others (but see Jimenez-Valverde et al. 2008), the more recent trend is to combine projections from different models in a so called "ensemble forecasting" approach (Araújo and New 2007) rather than trying to indentify a "best" model. Some of the results presented in this thesis tend to confirm that ensemble forecasting is indeed a promising approach (see Chapters 1.3 and 3.1).

Projection and evaluation of the model

Model evaluation is generally done by comparing the models' predictions with some, ideally independent, validation data. In the case where presence and absence data are available for evaluation, widely used evaluation measures are the area under the relative operating characteristic curve (AUC, Hanley and McNeil 1982), Cohen's Kappa (Cohen 1969, Monserud and Leemans 1992), and the true skill statistic (TSS, Allouche et al. 2006). Because the Kappa and the TSS measures depend on a reclassification threshold (see last paragraph of this section), they are generally computed over the full range of possible thresholds and the best value is kept.

When absence data are unavailable, one possibility is to evaluate how much the predictions differ from randomness (Boyce et al. 2002, and see also continuous Boyce index in Hirzel et al. 2006). E.g., a low habitat suitability class should contain less occurrence

records than expected by chance, whereas a high habitat suitability class should contain more of them. Randomness is when the number of presences found in each suitability class is proportional to the surface covered by this class.

Ideally, a model should always be evaluated against completely independent data (Araújo et al. 2005). In practice, however, independent data are at best scarce and most of the time completely absent. For this reason, most studies using SDMs base their model evaluation on "pseudo-independent" validation procedures, such as data-splitting procedures (e.g. cross-validation, split-sampling or bootstrapping) in which a part of the data are used for model calibration (e.g. 70%) and the remainder is kept for evaluation (e.g. 30%). An interesting implementation of such data-splitting is provided by the BIOMOD package (Thuiller et al. 2009 – Annex 2), where the random splitting can be repeated a user-defined number of times and the results indicating model performance are averaged.

The final aim when calibrating a SDM is generally to project it over a larger geographical area than the sampled locations. SDMs are most safely projected in the geographic area where the data they were calibrated with come from (interpolation). But they have also been transposed to different geographical locations (transposition in space, e.g. Randin et al. 2006) or into reconstructed past (e.g. Pearman et al. 2008b – Annex 4), and simulated future environmental conditions (e.g. Thomas et al. 2004). Such transposition in space and/or time involves further assumptions and limitations that I discuss in a next section of this introduction.

For practical reasons, it is often necessary to reclassify the raw projections of SDMs that are of a continuous nature (generally in the range 0-1) into binary projections that represent either suitable (potential presence of the species) or unsuitable (potential absence of the species) habitat. There are a number of approaches available and these are reviewed and assessed in Liu et al. (2005). To give only a few examples, one can chose the reclassification threshold as to reflect the prevalence of the absences and presences in the calibration data, chose it as to maximize the Kappa statistic (e.g. Engler et al. 2004 – Chapter 1.1) or the TSS (e.g. Chapter 3.1), or use a sensitivity-specificity sum maximization approach (Cantor et al. 1999, Manel et al. 2001; where sensitivity = proportion of correctly predicted presences and specificity = proportion of correctly predicted absences). A comparison of methods can be found in Liu et al. (2005) who recommend using the sensitivity-specificity sum maximization or the prevalence approach. Fixed thresholds and Kappa-maximizing thresholds are recommended against.

Overcoming practical limitations to the use of SDMs: how to deal with scarce or missing data?

One of the strengths of species distribution models (SDMs) compared to their more mechanistic counterparts such as dynamic global vegetation models (DGVMs) or gap models is their comparatively low data requirement. Nevertheless, even for SDMs, one can often be in a situation where data availability and/or quality is a limiting factor.

Data on species presence, absence and/or abundance can sometimes be collected through remote sensing (e.g. coastal vegetation mapping; Phinn et al. 2000) or aerial surveys (e.g. kangaroo monitoring; Pople et al. 2007), but for the large majority of species, such data can only be collected through field-work. While field-work has many positive aspects, starting

with the possibility it offers to actually observe the target species in its habitat, it remains a time consuming activity and has thus to be planed carefully. For this reason, using already existing information on species occurrence, as found for instance in herbaria and natural history collections (see Graham et al. 2004 for a review on these data sources) is an attractive alternative, and sometimes the only that is economically realistic.

Unfortunately, such data generally have several limitations (Graham et al. 2004): first and maybe most importantly, this type of data is typically characterized by a lack or reliable species absence information, i.e. georeferenced records attesting the absence of a species. Second, these data have often poor, unknown and/or variable spatial accuracy, and third, such data have generally been collected without any rigorous sampling scheme (e.g. they are biased towards easily accessible locations) and thus often contain some bias that can be problematic when trying to infer parameters from them later on (Yoccoz et al. 2001) and result in lower quality models (Edwards et al. 2006).

Among these issues, the lack of reliable absence information is so common that it has received considerable attention. One option to accommodate the lack of absence records is to use a modeling methodology that requires only presence data (e.g. environmental envelopes, ENFA, MAXENT; see Table 1). However, a potential drawback of these techniques is that they tend to overpredict the actual distribution of species, and some have suggested that their projections tend to be more towards the fundamental end of the fundamental vs. realized niche gradient than those of models calibrated on both presence and absence data (Zaniewski et al. 2002, Jimenez-Valverde et al. 2008). A second option that has been proposed to accommodate the lack of true absence information has been to artificially generate them, the so called "pseudo-absences" (Ferrier and Watson 1997 *in* Zaniewski et al. 2002). Different strategies for the generation of these pseudo-absences have been tested (Zaniewski et al. 2002, Engler et al. 2004 – Chapter 1.1, Zarnetske et al. 2007, Phillips et al. 2009) and comparisons between strategies have been carried-out (Chefaoui and Lobo 2008).

When working with rare species, all the above mentioned difficulties are generally exacerbated: one might not only lack absence records, but simply lack any occurrence data or have very few of them. An original solution to this problem is presented by Edwards et al. (2005) who, to model the distribution of rare lichens, modeled the distribution of another, more common, lichen species associated with it. However data about such association is not always available and in this case the only option remains to collect more data in the field. Unfortunately, even if time and funding is available for carrying-out field work, using traditional sampling methods can prove highly ineffective when trying to obtain new occurrence data for rare species (Yoccoz et al. 2001, Rushton et al. 2004). The use of SDMs to direct the sampling towards areas of high habitat suitability for the target species, thereby increasing the likelihood of finding new populations, is therefore a promising approach (e.g. Edwards et al. 2005).

Finally, having data with varying spatial accuracy highlights the dilemma of data quantity (number of occurrences) vs. quality (locational accuracy). In other words, given a pool of occurrences with variable spatial accuracy (i.e. some records are precisely georeferenced and others less), should one use all the data and sacrifice spatial accuracy (the projected model will have the grain size of the least accurate record), or keep only the most accurate records at the cost of having less data (i.e. the lesser accurate data will have to be discarded)? Or maybe the solution to this dilemma lies in not solving it: calibrating different models at different resolutions, and combine their projections?

Overcoming Methodological limitations of SDMs: From static to dynamic species distribution models.

Besides the wide range of applications in conservation biology, an increasingly popular domain where species distribution models (SDMs) are employed is to project the distribution of species under forecasted global change scenarios (e.g. Guisan and Theurillat 2000b, Bakkenes et al. 2002, Thuiller et al. 2005). Such projections can then be, for instance, used to estimate the potential threat of climate change to biodiversity (e.g. Thomas et al. 2004).

However, whether they are based on simple bioclimatic envelopes, logistic regression or more complex algorithms, classical implementations of SDMs suffer from a number of limitations that become particularly relevant when these models are used for projecting distributions of species under future climatic conditions. These limitations mainly arise from the "static" nature of SDMs and are (Pearson and Dawson 2003, Pearman et al. 2008a): (i) the assumption of niche conservatism (i.e. potential changes in biotic interactions and phenotypic and evolutionary adaptations are ignored) and, (ii) the absence of dispersal implementation.

One reason underlying the need for the assumption of niche conservatism is that SDMs implicitly account for biotic interactions and have a static view of these relationships. While accounting implicitly for competition and other biotic interactions is often a desirable propriety because it greatly simplifies model calibration, it might also lead to misleading projections if the interactions between species are altered, and this has been shown to possibly occur under climate change conditions (Davis et al. 1998a, Suttle et al. 2007). However, the severity of this limitation certainly depends on the type of species that is modeled: Pearman et al. (2008b – Annex 4) illustrate for instance that dominant species show greater conservatism in their realized niche than others. Furthermore, Pearson and Dawson (2003) also argue that, at large scales, biotic interactions become unimportant in driving a species distribution, and that they are thus suitable for making broad projections of potential climate change impacts (but see Araújo and Luoto 2007, who find contradictory results to this widely held view).

Another reason why the niche conservatism assumption is needed is because SDMs are unable to forecast how species might genetically adapt or express phenotypic plasticity to changing environmental conditions (Theurillat and Guisan 2001). While some studies have shown evidence of rapid evolutionary change (Thomas et al. 2001), others have found niche conservatism over several millions of years (Peterson et al. 1999, see Pearman at al. 2008a for a review) or rates of evolutionary change much slower than required for species to evolve necessary adaptations in the face of climate change (Davis and Shaw 2001, Etterson and Shaw 2001). Such contradictory evidence again suggests that niche conservatism depends on the considered species. E.g., long lived and low-dispersing species are thought to have higher conservatism of their realized niche (Pearson and Dawson 2003).

The second major methodological limitation of SDMs is that they do not take species dispersal and population dynamics into account. Again, this is not necessarily an important issue when projections are carried out under current climatic conditions (assuming the target species has reached its environmental equilibrium state), but becomes a potentially serious problem when projections are carried-out for future climatic conditions. The reason for that is obvious: the projections yielded by SDMs indicate the locations expected to become potentially suitable for the species in the future, but they are not concerned about whether the species will actually be able to physically disperse in order to reach these newly suitable locations. This shortcoming is further exacerbated by the nowadays increasing landscape fragmentation that hampers species dispersal (Pitelka et al. 1997).

In the absence of dispersal limitation implementation, studies that have projected potential impact of climate change on plant species have considered two extreme scenarios: "no dispersal" and "unlimited dispersal" (e.g. Thomas et al. 2004, Thuiller et al. 2005). Under the pessimistic "no-dispersal" assumption, a species is considered to have no dispersal capacity and thus it is projected to persist only if those habitats that it already occupies remain suitable under future climatic conditions. At the optimistic end, the "unlimited dispersal" (sometimes also called "universal dispersal", Thomas et al. 2004) hypothesis assumes that all suitable habitats will effectively become colonized.

However, these two extreme scenarios can yield large discrepancies in projections. For instance in a world-wide assessment (20% of the earth's emerged surface), Thomas et al. (2004) project species extinctions between 21-32% under "no-dispersal" and 38-52% under "unlimited" dispersal scenario. Although the exactitude of these numbers have been questioned (Buckley and Roughgarden 2004, Harte et al. 2004, Thuiller et al. 2004), they nevertheless illustrates the large degree of incertitude that is associated with such projections in the absence of dispersal implementation.

Similarly, in an assessment of the potential threats of climate change to 1'350 European plant species by 2080 (Thuiller et al. 2005), the difference in rates of species classified as critically endangered between "unlimited" and "no-dispersal" scenario is more than double (~3-7% under unlimited dispersal vs. ~10-25% under no-dispersal). These important discrepancies between unlimited and no-dispersal scenarios highlight the need for a more dynamic implementation of SDMs so that they can account for dispersal limitations.

A "static" and "low data requirement" vs. "dynamic" and "data hungry" gradient

As was briefly mentioned in the first section of this introduction, another approach for modeling the distribution of species, and in particular vegetation, is to use mechanistic models that are generally referred to as "dynamic global vegetation models" (DGVM, see Woodward and Lomas 2004 for a review, and Cramer et al. 2001 for a comparison of six different DGVMs).

Unlike SDMs that are correlative in nature, DGVMs are more or less completely mechanistic (process-based). They attempt to model the different physiological, biophysical and biogeochemical processes that are involved in vegetation growth, by mechanistically modeling elements such as photosynthesis, respiration, biomass and nutrient cycling such as the allocation of carbon and nitrogen within the plant (e.g. François et al. 2006). DGVMs also model elements of vegetation dynamics, such as competition, establishment and mortality. As a result, DGVMs avoid some the above mentioned problems encountered by SDMs when projected under future climatic conditions, and this has led some to consider them as only viable option (Woodward and Beerling 1997, Woodward and Lomas 2004).

However, as noted by Pearson and Dawson (2003) who compare SDMs and DGVMs in a review paper, DGVMs have also their shortcomings. To begin with, they are no more able than SDMs to account for evolutionary changes that would affect a species fundamental niche. Furthermore, even in the case where they correctly forecast the future distribution of a species' fundamental niche, it is likely that this will not match the species actual distribution (i.e. realized niche), which is ultimately the object of interest for practical applications. Finally, the higher complexity of DGVMs also comes at a price: their calibration requires much more data and more diverse data than that of SDMs (this is here referred to as being "data hungry"). In practice, this means that DGVMs, in their current state of development, are only able to model broad categories of vegetation (usually referred to as "plant functional types") and do so

only at a coarse spatial resolution. Modeling individual species, not to mention a large number of individual species, is thus typically not possible using DGVMs, and, for this kind of application, SDMs currently remain the only option.

In Guisan and Zimmermann (2000), SDMs and DGVMs are presented as two different offerings in a trade-off between precision, at the SDM end of the spectrum, and generality, at the DGVM end of the spectrum. Here I propose a slightly modified version of this gradient, with a "correlative", "static", "low data requirement" end of the spectrum where we find SDMs, and a "mechanistic", "dynamic", "data hungry" end where we find DGVMs (Figure 2).

Given that both classical SDMs and DGVMs have their strength and weaknesses, the idea of combining them to get the best of both worlds is an appealing option and a number of studies have explored this possibility (e.g., Carrey 1996, Dullinger et al. 2004). The idea developed by the majority of these studies is to use classical SDMs to project a species' habitat suitability (i.e. potential distribution), and then model its actual distribution by restricting it through a number of mechanistic dispersal and population dynamics parameters. Typically, the colonization probability of an empty grid cell (i.e. sink cell) will depend on the distance, the number and the fecundity of surrounding occupied cells (i.e. source cells). While this approach does not solve the problems related to niche conservatism, it at least improves on aspects related to dispersal and population dynamics.

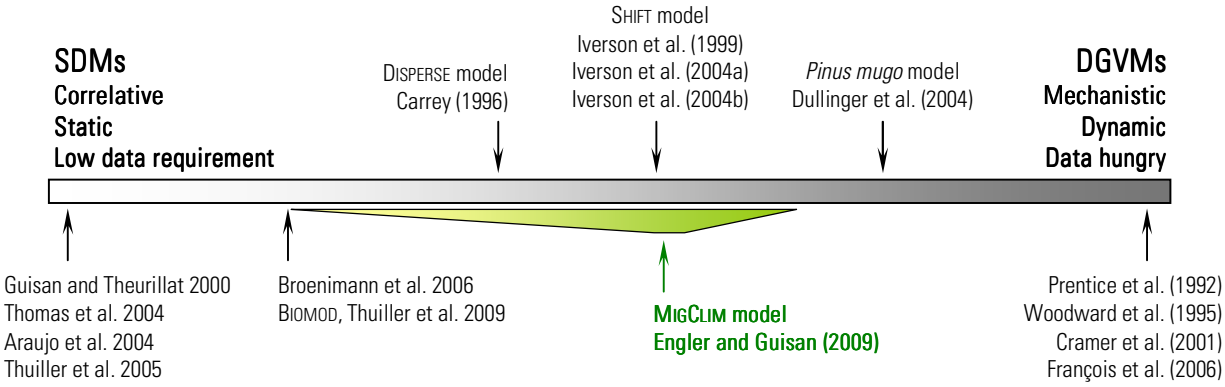


Figure 2. Examples of studies that have projected impacts of climate change on vegetation, classified along a gradient of how "static" (i.e. correlative; to the left) or "dynamic" (i.e. mechanistic; to the right) they are. Classification is semi-quantitative. The polygon indicates that the MIGCLIM model is flexible and its position along the gradient thus depends on the type of implementation that is chosen by the user.

The MIGCLIM model

Even though the need for including dispersal limitations into models has been raised repeatedly (Pitelka et al. 1997, Woodward and Beerling 1997, Cain et al. 1998, Davis et al. 1998b, Nathan and Muller-Landau 2000, Ronce 2001, Araujo and Guisan 2006, Midgley et al. 2007, Thuiller et al. 2008), so far relatively few niche-based climate change modeling studies have considered it (e.g. Carey 1996, Ostendorf et al. 2001, Dullinger et al. 2004, Iverson et al. 1999, Iverson et al. 2004a). Furthermore, none of the developed models so far seems to be general and user friendly enough to be implemented easily outside of the context it has been

designed for (except BIOMOD; Thuiller et al. 2009 – Annex 2, but it allows implementing of dispersal only to a limited extent).

The idea behind the development of MIGCLIM was to provide a good balance between model complexity, generality and usability. Practically, this means that I developed the MIGCLIM model to be flexible in its parameters, so that it can be adjusted to cover a wider range of the "static" vs. "dynamic" gradient (Figure 2). In its simplest form, MIGCLIM can be used to simply restrict the dispersal of a species by a fixed distance, achieving a result similar to what was done by Broenimann et al. (2006). In its more complex form, MIGCLIM can account for a large number of different parameters such as seed dispersal kernel, evolution of a population's reproductive potential over time, stochastic long distance dispersal events, barriers to dispersal, random population extinctions, vegetative and seed bank resilience to environmental change or differential dispersal along river or road features, e.g. to simulate a species that would preferentially disperse along rivers or that is thought to be frequently transported over large distances along roads. The full list of model parameters and explanations can be found in the MIGCLIM user guide given in Annex 5 of this thesis. Obviously, all degrees of complexity between these two extremes can be used and it ultimately depends oh how much data the user possesses to calibrate the parameters and how much confidence these data can be awarded.

Assessment of potential climate change impact using fine scale data

Ever increasing observations that the anthropogenic climate change (IPCC 2007) is affecting spatial distribution of plant and animal species around the world (Parmesan and Yohe 2003, Walther et al. 2005, Lenoir et al. 2008, Parolo and Rossi 2008) have triggered growing interest to forecast the potential impacts of the forecasted climate change on species distributions. Furthermore, given that the political decisions needed to mitigate the extent and impact of global warming are mostly of unpopular nature (e.g. raising taxes on fossil fuels), scientific evidence and assessment of how much climate change already has and will potentially impact on our environment are valuable.

In this context, and despite their limitations (discussed in the previous section of this introduction), species distribution models (SDMs) have been used for instance to assess how climate change is expected to affect biodiversity (Thomas et al. 2004, Thuiller et al. 2005), but also the distribution of disease transmitting species that jeopardize human health (Peterson and Shaw 2003, Lopez-Cardenas et al. 2005) or the invasive potential of alien species (Bradley et al. 2009).

Among the studies that assessed the potential impacts of the forecasted climate change on plant and animal species using SDMs, two groups can clearly be distinguished: those that were carried out over a large geographical extent but at a coarse spatial resolution (i.e. large grain size) and those that were conducted with high spatial accuracy (i.e. small grain size) but over a fairly limited geographical extent.

In the first group, which we will here term as "global scale", studies were typically carried-out at a continental scale (e.g. Huntley et al. 1995, Bakkenes et al. 2002, Thuiller et al. 2005) or world-wide (e.g. Thomas et al. 2004), with a spatial resolution of ~10-50 km, sometimes a bit less (e.g. Iverson et al. 2004a apply a model with 1 km pixel size over half of the united states). A typical example of this category is the work by Thuiller et al. (2005) who

calibrated Europe-wide models for 1345 species at a resolution of 50 km and projected them at a resolution of ~15 km.

The studies of the second group, that we will term as "local scale", were carried-out over a much smaller spatial extent, generally a few hundred or a few thousand square kilometers, but with a spatial resolution typically in the range of 20-100 m. Examples of this category includes the works form Guisan and Theurillat (2000b), Dirnbök et al. (2003) or Randin et al. (2006).

While global scale studies have by definition been carried out over large proportion of the planet, this is not the case of local scale studies that are generally restricted to one or possibly two study areas of limited spatial extent. Furthermore, because different studies often use different methodologies and climate change scenarios, the results can be difficult to compare. There is thus a need for conducting assessments of potential climate change impacts at a fine scale but over a large geographic extent and with a consistent methodology.

Objectives and structure of the thesis

In this thesis, my objective is to contribute to the development of methodological improvement in the field of species distribution modeling. More precisely, the objectives are to propose solutions to overcome limitations of species distribution models when applied to conservation issues (Chapter 1) or used as an assessment tool for potential impacts of global change (Chapters 2 and 3).

Chapter 1 – **Improving modeling and sampling strategies for rare and endangered species** – presents a methodology to generate pseudo-absences, assesses the data quality vs. quantity dilemma and illustrates how SDMs can be successfully used to model and sample rare species (Section 1.1 – Engler et al. 2004). This chapter also provides a comprehensive testing of the model-based sampling approach, with both theoretical (simulation-based; Section 1.2) and practical (field-based; Section 1.3) validation.

In **Chapter 2** – **Moving forwards: from static to dynamic species distribution models** – I develop a dynamic model, MIGCLIM, that allows the implementation of dispersal limitations into SDMs and present an application of this model to two virtual species (Section 2.1 – Engler and Guisan 2009). Given that accounting for dispersal limitations requires information on seed dispersal distances, a general methodology to classify species into broad dispersal types is also developed (Section 2.2 – Vittoz and Engler 2007). Finally, the MIGCLIM model is applied to a large number of species in a study area of the western Swiss Alps (Section 2.3 – Engler et al. 2009). This allows to evaluate if accounting for dispersal limitations has a significant influence on forecasts of species threat levels from climate change.

Chapter 3 – **Assessment of potential climate change impacts using fine scale data** – presents the largest fine scale assessment of potential impacts of climate change on mountain vegetation that has been carried-out to date (section 3.1). This assessment involves vegetation from 12 study areas distributed across all major European mountain ranges.

Besides the publications presented in chapters 1-3, that I have led or to which I provided a major contribution, my research interests have also offered me the opportunity to make

collaborations on other projects with subjects closely related to the work presented in chapters 1-3. The results of some of these collaborations that are most relevant to the material presented in the main body of this thesis are given in **Annexes 1-4**. Finally, **Annex 5** is a user guide for the MIGCLIM tool that I developed.

Annex 1 – **Using niche-based models to improve the sampling of rare species**, Guisan et al. (2006) – is the full text from which chapter 1.2 is an abstract. This publication presents a comprehensive testing of the model-based sampling methodology through simulations (same as what is presented in chapter 1.2) and field-work based validation.

Annex 2 – **BIOMOD, a platform for ensemble forecasting of species distributions**, Thuiller et al. (2009) – presents a free and open source modeling platform that allows model calibration for multiple species and with multiple methods. I made a modest contribution to this modeling package by implementing a technique-independent method to derive the contribution of each predictor in a given model. This is also the package that I have used in chapter 3.1 to calibrate and project distribution models under current and future climatic conditions.

Annex 3 – **Climate change and plant distribution: local models predict high-elevation persistence**, Randin et al. (2009) – This publication is based on the data from two of the study areas that are also used in chapter 3.1 (Swiss Western and Swiss inner Alps) and compares the results of coarse scale and fine scale models when used to project species distributions for future climatic conditions. It shows that local scale models can predict significantly more species persistence than global scale models and thus that more fine scale assessment of climate change impacts, as presented in chapter 3.1, are needed.

Annex 4 – **Prediction of plant species distributions across six millennia**, Pearman et al. (2008) – presents a study of so called "hindcasting" of species distribution models. The idea is that, while we are unable verify model predictions for the future, projections of species distribution in the past can be validated through confrontation with distributions reconstructed from palynological records.

Annex 5 – **MIGCLIM User Guide** – is a user guide to the MIGCLIM tool that is central to the chapter 2 of this thesis.

References

Acosta L.E., 2008. Distribution of Geraeocormobius sylvarum (Opiliones, Gonyleptidae): Range modeling based on bioclimatic variables. *The Journal of Arachnology*, **36**, 574–582.

Allouche O., Tsoar A. and Kadmon R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**(6), 1223-1232.

Araujo M.B. and Guisan A., 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**(10), 1677-1688.

Araujo M.B. and Luoto M., 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**(6), 743-753.

Araújo M.B., Cabeza M., Thuiller W., Hannah L. and Williams P.H., 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology*, **10**, 1–9.

Araújo M.B., Pearson R.G., Thuiller W. and Erhard M., 2005. Validation of species-climate impact models under climate change. *Global Change Biology*, **11**(9), 1504-1513.

Araújo M.B. and New M., 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**(1), 42-47.

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**(6), 1541-1550.

Austin M.P. and Meyers J.A., 1996. Current approaches to modelling the environmental niche of eucalypts: Implication for management of forest biodiversity. *Forest Ecology and Management*, **85**(1-3), 95-106.

Bakkenes M., Alkemade J.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**(4), 390-407.

Botkin D.B., Janak J.F. and Wallis J.R., 1972. Some ecological consequences of a computer model of forest growth. *The Journal of Ecology*, **60**, 849-872.

Box E.O. 1981. Macroclimate and Plant Forms: An Introduction to Predictive Modeling in Phytogeography. - Junk.

Boyce M.S., Vernier P.R., Nielsen S.E. and Schmiegelow F.K.A., 2002. Evaluating resource selection functions. *Ecological Modelling*, **157**(2-3), 281-300.

Bradley B.A., Oppenheimer M. and Wilcove D.S., 2009. Climate change and plant invasions: restoration opportunities ahead? *Global Change Biology*, **15**, 1511–1521.

Breiman L., Friedman J.H., Olshen R.A. and Stone C.J. 1984. Classification and Regression Trees. - Chapman & Hall.

Breiman L., 2001. Random Forests. *Machine Learning*, **45**(1), 5-32.

Broennimann O., Thuiller W., Hughes G., Midgley G.F., Alkemade J.M.R. and Guisan A., 2006. Do geographic distribution, niche property and life form explain plants' vulnerability to global change? *Global Change Biology*, **12**(6), 1079-1093.

Broennimann O., Treier U.A., Müller-Schärer H., Thuiller W., Peterson A.T. and Guisan A., 2007. Evidence of climatic niche shift during biological invasion. *Ecology Letters*, **10**(8), 701-709.

Brown J.H., Stevens G.C. and Kaufman D.M., 1996. The geographic range: Size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, **27**, 597-623.

Brown J.L., Morales V. and Summers K., 2009. Home range size and location in relation to reproductive resources in poison frogs (Dendrobatidae): a Monte Carlo approach using GIS data. *Animal Behaviour*, **77**, 547–554.

Brzeziecki B., Kienast F. and Wildi O., 1995. Modeling Potential Impacts of Climate-Change on the Spatial-Distribution of Zonal Forest Communities in Switzerland. *Journal of Vegetation Science*, **6**(2), 257-268.

Buckley L.B. and Roughgarden J., 2004. Biodiversity conservation - Effects of changes in climate and land use. *Nature*, **430**(6995).

Busby J.R. 1991. BIOCLIM - A bioclimate analysis and prediction system. - In: C.R. Margules and Austin, M.P. (eds.), Nature conservation: cost effective biological surveys and data analysis. CSIRO, pp. 64-68.

Cain M.L., Damman H. and Muir A., 1998. Seed Dispersal and the Holocene Migration of Woodland Herbs. *Ecological Monographs*, **68**(3), 325-347.

Cantor S.B., Sun C.C., Tortolero-Luna G., Richards-Kortum R. and Follen M., 1999. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. *Journal of Clinical Epidemiology*, **52**, 885-892.

Carey P.D., 1996. DISPERSE: A cellular automaton for predicting the distribution of species in a changed climate. *Global Ecology and Biogeography Letters*, **5**(4-5), 217-226.

Chase J.M. and Leibold M.A., 2003. Ecological Niches: Linking Classical and Contemporary Approaches. The University of Chicago Press, Chicago, IL, USA.

Chefaoui R.M. and Lobo J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, **210**(4), 478-486.

Cohen J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**(1), 37-46.

Cramer W., Bondeau A., Woodward F.I., Prentice I.C., Betts R.A., Brovkin V., Cox P.M., Fisher V., Foley J.A., Friend A.D., Kucharik C., Lomas M.R., Ramankutty N., Sitch S., Smith B., White A. and 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**(4), 357-373.

Davis A.J., Jenkinson L.S., Lawton J.H., Shorrocks B. and Wood S., 1998a. Making mistakes when predicting shifts in species range in response to global warming. *Nature*, **391**(6669), 783-786.

Davis A.J., Lawton J.H., Shorrocks B. and Jenkinson L.S., 1998b. Individualistic species response invalidate simple physiological model of community dynamics under global environmental change. *Journal of Animal Ecology*, **67**, 600-612.

Davis M.B. and Shaw R.G., 2001. Range Shifts and Adaptive Responses to Quaternary Climate Change. *Science*, **292**, 673-679.

Dirnböck T., Dullinger S. and Grabherr G., 2003. A regional impact assessment of climate and land-use change on alpine vegetation. *Journal of Biogeography*, **30**(3), 401-417.

Dullinger S., Dirnbock T. and Grabherr G., 2004. Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. *Journal of Ecology*, **92**(2), 241-252.

Edwards T.C., Cutler D.R., Zimmermann N.E., Geiser L. and Alegria J., 2005. Model-based stratifications for enhancing the detection of rare ecoligical events. *Ecology*, **86**(5), 1081-1090.

Edwards T.C., Cutler D.R., Zimmermann N.E., Geiser L. and Moisen G.G., 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecological Modelling*, **199**(2), 132-141.

Elith J., Graham C.H., Anderson R.P., Dudik M., Ferrier S., Guisan A., Hijmans R.J., Huettmann F., Leathwick J.R., Lehmann A., Li J., Lohmann L.G., Loiselle B.A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J.M., Peterson A.T., Phillips S.J., Richardson K., Scachetti-Pereira R., Schapire R.E., Soberon J., Williams S., Wisz M.S. and Zimmermann N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**(2), 129-151.

Elith J. and Graham C.H., 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32**, 66-77.

Engler R., Guisan A. and Rechsteiner L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41**(2), 263-274.

Engler R. and Guisan A., 2009. MIGCLIM: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, **15**(4), 590–601.

Estrada-Pena A. and Venzal J.M., 2007. Climate niches of tick species in the Mediterranean region: Modeling of occurrence data, distributional constraints, and impact of climate change. *Journal of Medical Entomology*, **44**(6), 1130-1138.

Etterson J.R. and Shaw R.G., 2001. Constraint to Adaptive Evolution in Response to Global Warming. *Science*, **294**, 151-154.

Ferrier S. and Watson G. 1997. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. - Environment Australia.

Ferrier S., Drielsma M., Manion G. and Watson G., 2002a. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation*, **11**(12), 2275-2307.

Ferrier S., Drielsma M., Manion G. and Watson G., 2002b. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation*, **11**(12), 2309-2338.

François L., Ghislain M., Dominique O. and Micheels A., 2006. Late Miocene vegetation reconstruction with the CARAIB model. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **238**, 302–320.

Franklin J., 1995. Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**(4), 474-499.

Friedman J.H., 1991. Multivariate Adaptive Regression Splines. *Annals of Statistics*, **19**(1), 1-67.

Galparsoro I., Borja A., Bald J., Liria P. and Chust G., 2009. Predicting suitable habitat for the European lobster (Homarus gammarus), on the Basque continental shelf (Bay of Biscay), using Ecological-Niche Factor Analysis. *Ecological Modelling*, **220**, 556–567.

Graham C.H., Ferrier S., Huettman F., Moritz C. and Peterson A.T., 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, **19**(9), 497-503.

Guisan A. and Theurillat J.P., 2000a. Equilibrium modeling of alpine plant distribution: how far can we go? *Phytocoenologia*, **30**(3-4), 353-384.

Guisan A. and Theurillat J.-P., 2000b. Assessing alpine plant vulnerability to climate change: A modeling perspective. *Integrated Assessment*, **1**, 307-320.

Guisan A. and Zimmermann N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**(2-3), 147-186.

Guisan A., Edwards T.C. and Hastie T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**(2-3), 89-100.

Guisan A. and Hofer U., 2003. Predicting reptile distributions at the mesoscale: relation to climate and topography. *Journal of Biogeography*, **30**(8), 1233-1243.

Guisan A. and Thuiller W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Guisan A., Broennimann O., Engler R., Vust M., Yoccoz N.G., Lehmann A. and Zimmermann N.E., 2006. Using niche-based models to improve the sampling of rare species. *Conservation Biology*, **20**(2), 501-511.

Guisan A., Zimmermann N.E., Elith J., Graham C., Phillips S. and Peterson A.T., 2007. What matters for predicting spatial distributions of trees: techniques, data, or species' characteristics? *Ecological Monographs*, **77**(4), 615-630.

Hanley J.A. and McNeil B.J., 1982. The Meaning and Use of the Area under a Receiver Operating Characteristic (Roc) Curve. *Radiology*, **143**(1), 29-36.

Harte J., Ostling A., Green J.L. and Kinzig A., 2004. Biodiversity conservation - Climate change and extinction risk. *Nature*, **430**(6995).

Hastie T. and Tibshirani R., 1986. Generalized Additive Models. *Statistical Science*, **1**(3), 297-318.

Hastie T., Tibshirani R. and Buja A., 1994. Flexible Discriminant-Analysis by Optimal Scoring. *Journal of the American Statistical Association*, **89**(428), 1255-1270.

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. and Jarvis A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**(15), 1965-1978.

Hirzel A. and Le Lay G., 2008. Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, **45**(5), 1372-1381.

Hirzel A.H., Helfer V. and Metral F., 2001. Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, **145**(2-3), 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**(7), 2027-2036.

Hirzel A.H., Le Lay G., Helfer V., Randin C. and Guisan A., 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, **199**(2), 142-152.

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 surfaces. *Journal of Biogeography*, **22**, 967-1001.

Hutchinson G.E., 1957. Population Studies - Animal Ecology and Demography - Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415-427.

IPCC (ed.). 2007. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment. Report of the Intergovernmental Panel on Climate Change. - Cambridge University Press.

Iverson L. and Prasad A., 1998a. Estimating regional plant biodiversity with GIS modelling. *Diversity and Distributions*, **4**(2), 49-61.

Iverson L.R. and Prasad A.M., 1998b. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs*, **68**(4), 465-485.

Iverson L.R., Prasad A. and Schwartz M.W., 1999. Modeling potential future individual tree-species distributions in the eastern United States under a climate change scenario: a case study with Pinus virginiana. *Ecological Modelling*, **115**(1), 77-93.

Iverson L.R., Schwartz M.W. and Prasad A.M., 2004a. How fast and far might tree species migrate in the eastern United States due to climate change? *Global Ecology and Biogeography*, **13**(3), 209-219.

Iverson L.R., Schwartz M.W. and Prasad A.M., 2004b. Potential colonization of newly available tree-species habitat under climate change: an analysis for five eastern US species. *Landscape Ecology*, **19**(7), 787-799.

Jaberg C. and Guisan A., 2001. Modelling the distribution of bats in relation to landscape structure in a temperate mountain environment. *Journal of Applied Ecology*, **38**(6), 1169-1181.

Jimenez-Valverde A., Lobo J.M. and Hortal J., 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, **14**(6), 885-890.

Lassale G. and Rochard E., 2009. Impact of twenty-first century climate change on diadromous fish spread over Europe, North Africa and the Middle East. *Global Change Biology*, **15**, 1072–1089.

Lassueur T., Joost S. and Randin C.F., 2006. Very high resolution digital elevation models: Do they improve models of plant species distribution? *Ecological Modelling*, **198**(1-2), 139-153.

Lawler J.J. and Edwards T.C., 2002. Landscape patterns as habitat predictors: building and testing models for cavity-nesting birds in the Uinta Mountains of Utah, USA. *Landscape Ecology*, **17**(3), 233-245.

Leathwick J.R., 1995. Climatic Relationships of Some New-Zealand Forest Tree Species. *Journal of Vegetation Science*, **6**(2), 237-248.

Leathwick J.R., 2001. New Zealand's potential forest pattern as predicted from current species-environment relationships. *New Zealand Journal of Botany*, **39**(3), 447-464.

Lehmann A., 1998. GIS modeling of submerged macrophyte distribution using Generalized Additive Models. *Plant Ecology*, **139**(1), 113-124.

Lenoir J., Gégout J.C., Marquet P.A., De Ruffray P. and Brisse H., 2008. A Significant Upward Shift in Plant Species Optimum Elevation During the 20th Century. *Science*, **320**, 1768-1771.

Lischke H., Guisan A., Fischlin A. and Bugmann H. 1998a. Vegetation response to climate change in the Alps: modeling studies. - In: P. Cebon, Dahinden, U., Davies, H.C., Imboden, D. and Jaeger, C.C. (eds.), Views from the Alps: regional perspectives on climate change. MIT Press, pp. 309-350.

Lischke H., Loffler T.J. and Fischlin A., 1998b. Aggregation of individual trees and patches in forest succession models: Capturing variability with height structured, random, spatial distributions. *Theoretical Population Biology*, **54**(3), 213-226.

Liu C.R., Berry P.M., Dawson T.P. and Pearson R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**(3), 385-393.

Lopez-Cardenas J., Bravo F.E.G., Schettino P.M.S., Solorzano J.C.G., Barba E.R., Mendez J.M., Sanchez-Cordero V., Peterson A.T. and Ramsey J.M., 2005. Fine-scale predictions of distributions of Chagas disease vectors in the state of Guanajuato, Mexico. *Journal of Medical Entomology*, **42**(6), 1068-1081.

Luoto M., Toivonen T. and Heikkinen R.K., 2002. Prediction of total and rare plant species richness in agricultural landscapes from satellite images and topographic data. *Landscape Ecology*, **17**(3), 195-217.

Luoto M., Heikkinen R.K., Poyry J. and Saarinen K., 2006. Determinants of the biogeographical distribution of butterflies in boreal regions. *Journal of Biogeography*, **33**(10), 1764-1778.

Lutolf M., Kienast F. and Guisan A., 2006. The ghost of past species occurrence: improving species distribution models for presence-only data. *Journal of Applied Ecology*, **43**(4), 802-815.

Mac Nally R., Fleishman E., Fay J.P. and Murphy D.D., 2003. Modelling butterfly species richness using mesoscale environmental variables: model construction and validation for mountain ranges in the Great Basin of western North America. *Biological Conservation*, **110**(1), 21-31.

Maggini R., Guisan A. and Cherix D., 2002. A stratified approach for modeling the distribution of a threatened ant species in the Swiss National Park. *Biodiversity and Conservation*, **11**(12), 2117-2141.

Maggini R., Lehmann A., Zimmermann N.E. and Guisan A., 2006. Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography*, **33**(10), 1729-1749.

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**(5), 921-931.

Maxwell D.L., Stelzenmüller V., Eastwood P.D. and Rogers S.I., 2009. Modelling the spatial distribution of plaice (Pleuronectes platessa), sole (Solea solea) and thornback ray (Raja clavata) in UK waters for marine management and planning. *Journal of Sea Research*, **61**, 258–267.

McCullagh P. and Nelder J.A. 1989. Generalized Linear Models, 2nd edn. - Chapman & Hall, London, UK.

Meynard C.N. and Quinn J.F., 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, **34**(8), 1455-1469.

Midgley G.F., Thuiller W. and Higgins S.I. 2007. Plant species migration as a key uncertainty in predicting future impacts of climate change on ecosystems: progress and challenges. - In: J.G. Canadell, Pataki, D.E. and Pitelka, L.F. (eds.), Terrestrial ecosystems in a changing world. Springer-Verlag, pp. 129-137.

Mladenoff D.J., Sickley T.A., Haight R.G. and Wydeven A.P., 1995. A regional landscape analysis and prediction of favorable gray wolf habitat in the northern Great Lakes region. *Conservation Biology*, **9**, 279–294.

Monserud R.A. and Leemans R., 1992. Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, **62**, 275-293.

Mueller J., Bassler C., Stratz C., Klocking B. and Brandl R., 2009. Molluscs and climate warming in a low mountain range national park. *Malacologia*, **51**(1), 89-109.

Murase H., Nagashima H., Yonezaki S., Matsukura R. and Kitakado T., 2009. Application of a generalized additive model (GAM) to reveal relationships between environmental factors and distributions of pelagic fish and krill: a case study in Sendai Bay, Japan. *ICES Journal of Marine Science*, **66**, 1417–1424.

Nathan R. and Muller-Landau H.C., 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology and Evolution*, **15**, 278-285.

Ostendorf B., Hilbert D.W. and Hopkins M.S., 2001. The effect of climate change on tropical rainforest vegetation pattern. *Ecological Modelling*, **145**, 211-224.

Parmesan C. and Yohe G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**(6918), 37-42.

Parolo G. and Rossi G., 2008. Upward migration of vascular plants following a climate warming trend in the Alps. *Basic and Applied Ecology*, **9**, 100-107.

Pearman P.B., Guisan A., Broennimann O. and Randin C.F., 2008a. Niche dynamics in space and time. *Trends in Ecology and Evolution*, **23**(3), 149-158.

Pearman P.B., Randin C.F., Broennimann O., Vittoz P., van der Knaap W.O., Engler R., Le Lay G., Zimmermann N.E. and Guisan A., 2008b. Prediction of plant species distributions across six millennia. *Ecology Letters*, **11**(4), 357-369.

Pearson R.G. and Dawson T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**(5), 361-371.

Pellet J., Guisan A. and Perrin N., 2004. A concentric analysis of the impact of urbanization on the threatened European tree frog in an agricultural landscape. *Conservation Biology*, **18**(6), 1599-1606.

Peterson A.T., Soberon J. and Sanchez-Cordero V., 1999. Conservatism of ecological niches in evolutionary time. *Science*, **285**(5431), 1265-1267.

Peterson A.T., 2001. Predicting species' geographic distributions based on ecological niche modeling. *Condor*, **103**(3), 599-605.

Peterson A.T., 2003. Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology*, **78**(4), 419-433.

Peterson A.T. and Shaw J., 2003. Lutzomyia vectors for cutaneous leishmaniasis in Southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. *International Journal for Parasitology*, **33**(9), 919-931.

Peterson A.T., 2006a. Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics*, **3**, 59-72.

Peterson A.T., 2006b. Ecologic niche modeling and spatial patterns of disease transmission. *Emerging Infectious Diseases*, **12**(12), 1822-1826.

Peterson A.T., Lash R.R., Carroll D.S. and Johnson K.M., 2006. Geographic potential for outbreaks of Marburg hemorrhagic fever. *American Journal of Tropical Medicine and Hygiene*, **75**(1), 9-15.

Peterson A.T., Williams R. and Chen G., 2007. Modeled global invasive potential of Asian gypsy moths, Lymantria dispar. *Entomologia Experimentalis Et Applicata*, **125**(1), 39-44.

Phillips S.J., Anderson R.P. and Schapire R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**(3-4), 231-259.

Phillips S.J., Dudík M., Elith J., Graham C.H., Lehmann A., Leathwick J. and Ferrier S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**(1), 181-197.

Phinn S.R., Menges C., Hill G.J.E. and Stanford M., 2000. Optimizing Remotely Sensed Solutions for Monitoring, Modeling, and Managing Coastal Environments. *Remote Sensing of the Environment*, **73**(2), 117-132.

Pineda E. and Lobo J.M., 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, **78**, 182–190.

Pitelka L.F., Gardner R.H., Ash J., Berry S., Gitay H., Noble I.R., Saunders A., Bradshaw R.H.W., Brubaker L., Clark J.S., Davis M.B., Sugita S., Dyer J.M., Hengeveld R., Hope G., Huntley B., King G.A., Lavorel S., Mack R.N.,

Malanson G.P., McGlone M., Prentice I.C. and Rejmanek M., 1997. Plant migration and climate change. *American Scientist*, **85**(5), 464-473.

Pople A.R., Phinn S.R., Menke N., Grigg G.C., Possingham H.P. and McAlpine C., 2007. Spatial patterns of kangaroo density across the South Australian pastoral zone over 26 years: aggregation during drought and suggestions of long distance movement. *Journal of Applied Ecology*, **44**(5), 1068-1079.

Prentice I.C., Cramer W., Harrison S.P., Leemans R., Monserud R.A. and Solomon A.M., 1992. A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate. *Journal of Biogeography*, **19**(2), 117-134.

Pulliam H.R., 2000. On the relationship between niche and distribution. *Ecology Letters*, **3**(4), 349-361.

Qi D., Hu Y., Gu X., Li M. and Wei F., 2009. Ecological niche modeling of the sympatric giant and red pandas on a mountain-range scale. *Biodiversity Conservation*, **18**, 2127–2141.

Raes N., Roos M.C., Ferry Slik J.W., van Loon E. and der Steege H., 2009. Botanical richness and endemicity patterns of Borneo derived from species distribution models. *Ecography*, **32**, 180-192.

Randin C.F., Dirnbock T., Dullinger S., Zimmermann N.E., Zappa M. and Guisan A., 2006. Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**(10), 1689-1703.

Randin C.F., Engler R., Normand S., Zappa M., Zimmermann N.E., Pearman P.B., Vittoz P., Thuiller W. and Guisan A., 2009. Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology*, **15**, 1557-1569.

Randin C.F., Vittoz P., Jaccard H., Yoccoz N.G. and Guisan A., in press. Land use improves spatial predictions of plant abundances but not occurrences in a mountain landscape. *Journal of Vegetation Science*.

Raxworthy C.J., Martinez-Meyer E., Horning N., Nussbaum R.A., Schneider G.E., Ortega-Huerta M.A. and Peterson A.T., 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, **426**(6968), 837-841.

Raxworthy C.J., Ingram C.M., Rabibisoa N. and Pearson R.G., 2007. Applications of ecological niche modeling for species delimitation: A review and empirical evaluation using day geckos (Phelsuma) from Madagascar. *Systematic Biology*, **56**(6), 907-923.

Richardson B.J., Zabka M., Gray M.R. and Milledge G., 2006. Distributional patterns of jumping spiders (Araneae : Salticidae) in Australia. *Journal of Biogeography*, **33**(4), 707-719.

Rickebusch S., Lischke H., Bugmann H., Guisan A. and Zimmermann N.E., 2007. Understanding the low-temperature limitations to forest growth through calibration of a forest dynamics model with tree-ring data. *Forest Ecology and Management*, **246**(2-3), 251-263.

Ridgeway G., 1999. The state of boosting. *Computing Science and Statistics*, **31**, 172-181.

Ripley B.D., 1996. Pattern recognition and neural networks. Cambridge University Press.

Ronce O., 2001. Understanding plant dispersal and migration. *Trends in Ecology and Evolution*, **16**(12), 663.

Rushton S.P., Ormerod S.J. and Kerby G., 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology*, **41**(2), 193-200.

Segurado P. and Araujo M.B., 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**(10), 1555-1568.

Suttle K.B., Thomsen M.A. and Power M.E., 2007. Species interactions reverse grassland responses to changing climate. *Science*, **315**(5812), 640-642.

Svenning J.C. and Skov F., 2004. Limited filling of the potential range in European tree species. *Ecology Letters*, **7**(7), 565-573.

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 climates. *Journal of Biogeography*, **23**(2), 203-233.

Theurillat J.P. and Guisan A., 2001. Potential impact of climate change on vegetation in the European Alps: A review. *Climatic Change*, **50**(1-2), 77-109.

Thomas C.D., Bodsworth E.J., Wilson R.J., Simmons A.D., Davies Z.G., Musche M. and Conradt L., 2001. Ecological and evolutionary processes at expanding range margins. *Nature*, **411**(6837), 577-581.

Thomas C.D., Cameron A., Green R.E., Bakkenes M., Beaumont L.J., Collingham Y.C., Erasmus B.F.N., Ferreira de Siqueira M., Grainger A., Hannah L., Hughes L., Huntley B., Van Jaarsveld A.S., Midgley G.F., Miles L., Ortega-

Huerta M.A., Peterson A.T., Phillips O.L. and Williams S.E., 2004. Extinction risk from climate change. *Nature*, **427**(6970), 145-147.

Thuiller W., Araújo M.B., Pearson R.G., Whittaker R.J., Brotons L. and Lavorel S., 2004. Uncertainty in predictions of extinction risk. *Nature*, **430**, 34.

Thuiller W., Lavorel S., Araujo M.B., Sykes M.T. and Prentice I.C., 2005. Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(23), 8245-8250.

Thuiller W., Midgley G.F., Rouget M. and Cowling R.M., 2006. Predicting patterns of plant species richness in megadiverse South Africa. *Ecography*, **29**(5), 733-744.

Thuiller W., Albert C., Araujo M.B., Berry P.M., Guisan A., Hickler T., Midgley G.F., Paterson J., Schurr F.M., Sykes M.T. and Zimmermann N.E., 2008. Predicting global change impacts on plant species distributions: future challenges. *Perspective in Plant Ecology, Evolution and Systematics*, **9**, 137-152.

Thuiller W., Lafourcade B., Engler R. and Araújo M.B., 2009. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369-373.

Tittensor D.P., Baco A.R., Brewin P.E., Clark M.R., Consalvey M., Hall-Spencer J., Rowden A.A., Schlacher T., Stocks K.I. and Rodgers A.D., 2009. Predicting global habitat suitability for stony corals on seamounts. *Journal of Biogeography*, **36**, 1111–1128.

Tsoar A., Allouche O., Steinitz O., Rotem D. and Kadmon R., 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, **13**(4), 397-405.

Vittoz P. and Engler R., 2007. Seed dispersal distances: a typology based on dispersal modes and plant traits. *Botanica Helvetica*, **117**(2), 109-124.

Walther G.-R., Beissner S. and Burga C.A., 2005. Trends in the upward shift of alpine plants. *Journal of Vegetation Science*, **16**(5), 541-548.

Wollan A.K., Bakkestuen V., Kauserud H., Gulden G. and Halvorsen R., 2008. Modelling and predicting fungal distribution patterns using herbarium data. *Journal of Biogeography*, **35**, 2298–2310.

Woodward F.I., Smith T.M. and Emanuel W.R., 1995. A global primary productivity and phytogeography model. *Global Biogeochemical Cycles*, **9**, 471–490.

Woodward F.I. and Beerling D.J., 1997. The dynamics of vegetation change: health warnings for equilibrium 'dodo' models. *Global Ecology and Biogeography Letters*, **6**, 413–418.

Woodward F.I. and Lomas M.R., 2004. Vegetation dynamics - simulating responses to climatic change. *Biological Reviews*, **79**(3), 643-670.

Yoccoz N.G., Nichols J.D. and Boulinier T., 2001. Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution*, **16**(8), 446-453.

Zaniewski A.E., Lehmann A. and Overton J.M.C., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, **157**(2-3), 261-280.

Zarnetske P.L., Edwards T.C. and Moisen G.G., 2007. Habitat classification modeling with incomplete data: Pushing the habitat envelope. *Ecological Applications*, **17**(6), 1714-1726.

Chapter 1

Improving modeling and sampling strategies for rare and endangered species

1.1 An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data.

1.2 Testing the model-based sampling approach through simulation.

1.3 Using habitat - suitability models enhances chances to find rare species in the field.

1.1

An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data.

Robin Engler[1], Antoine Guisan[1] and Luca Rechsteiner[1]

This manuscript has been published in 2004 *Journal of Applied Ecology*, **41**, 263-274.

DOI: 10.1111/j.0021-8901.2004.00881.x

[1] Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland.

CONTRIBUTION TO THE PROJECT:
I developed the original idea in collaboration with LR and AG. I prepared the data, performed the modeling and field-work in collaboration with LR. I carried-out the analyses and simulations and wrote the initial draft of the manuscript. I also participated in the reviewing process.

Abstract

1. Few examples of habitat-modelling studies of rare and endangered species exist in the literature, although from a conservation perspective predicting their distribution would prove particularly useful. Paucity of data and lack of valid absences are the probable reasons for this shortcoming. Analytic solutions to accommodate the lack of absence include the ecological niche factor analysis (ENFA) and the use of generalized linear models (GLM) with simulated pseudo-absences.

2. In this study we tested a new approach to generating pseudo-absences, based on a preliminary ENFA habitat suitability (HS) map, for the endangered species *Eryngium alpinum*. This method of generating pseudo-absences was compared with two others: (i) use of a GLM with pseudo-absences generated totally at random, and (ii) use of an ENFA only.

3. The influence of two different spatial resolutions (i.e. grain) was also assessed for tackling the dilemma of quality (grain) vs. quantity (number of occurrences). Each combination of the three above-mentioned methods with the two grains generated a distinct HS map.

4. Four evaluation measures were used for comparing these HS maps: total deviance explained, best kappa, Gini coefficient and minimal predicted area (MPA). The last is a new evaluation criterion proposed in this study.

5. Results showed that (i) generalized linear models using ENFA-weighted pseudo-absence provide better results, except for the MPA value, and that (ii) quality (spatial resolution and locational accuracy) of the data appears to be more important than quantity (number of occurrences). Furthermore, the proposed MPA value is suggested as a useful measure of model evaluation when used to complement classical statistical measures.

6. *Synthesis and applications*. We suggest that the use of ENFA-weighted pseudoabsence is a possible way to enhance the quality of GLM-based potential distribution maps and that data quality (i.e. spatial resolution) prevails over quantity (i.e. number of data). Increased accuracy of potential distribution maps could help to define better suitable areas for species protection and reintroduction.

Keywords: ecological niche factor analysis (ENFA), *Eryngium alpinum*, generalized linear model (GLM), habitat suitability (HS) model, minimal predicted area (MPA), spatial resolution vs. data size.

Introduction

A variety of predictive models is currently in use to simulate the spatial distribution of plant and animal species (Franklin 1995; Guisan & Zimmermann 2000; Scott et al. 2002). Most of the models rely on adjusting a quantitative relationship between a taxon and its direct environment. Models have been developed for plant communities (Brzeziecki, Kienast & Wildi 1993; Brown 1994; Zimmermann & Kienast 1999), for individual plant species (Guisan, Theurillat & Kienast 1998; Guisan, Weiss & Weiss 1999; Peterson 2001; Bakkenes et al. 2002) and for plant species assemblages and biodiversity reconstructed from superimposing individual species' predictions (Guisan & Theurillat 2000; Lehmann, Overton & Leathwick 2002).

These models result in spatial predictions indicating locations of the most suitable (and unsuitable) habitats for a target species, community or biodiversity (i.e. indicating 'hotspots'). Generalized linear and generalized additive models (GLM and GAM), implemented within a geographical information system (GIS), have become very popular for predicting such distributions (Guisan, Edwards & Hastie 2002).

However, as yet, relatively few predictive models have been applied to rare and endangered species (Miller 1986; Myatt 1987; Carey & Brown 1995; Godown & Peterson 2000; Elith & Burgman 2002), despite their potential in conservation management, for instance in identifying sites with high potential for colonization. This may be because (i) data for rare and endangered taxa very often consist of a set of observed occurrences without sites of observed absences (hereafter called presence-only data); (ii) data for a single taxon are usually scarce (few observations); and (iii) often observations are not associated with any defined sampling unit (of known surface area) or they lack sufficient locational accuracy.

The first problem is commonly associated with data stored in large biological data bases. Such data have often been recorded by volunteers, usually without recourse to any predefined sampling strategy. Scarcity of data is specific to uncommon and rare species, for which prevalence in a data bank is, by definition, very low. Historical records, such as herbarium or museum collections, often lack precise details of location: at best they show proximity to a common site, a valley or village at a scale of a kilometre or more. These two problems make it more difficult to apply the usual statistical approaches. Such data contrast unfavourably with recent observations (≤ 10 years) sampled using a global positioning system (GPS) with a much higher spatial accuracy.

This highlights the dilemma of quantity (number of occurrences) vs. quality (locational accuracy). When the spatial accuracy associated with the geographical location of each observation site is known (e.g. the true site has a 95% probability of being within a 100-m radius), it becomes a major consideration in choosing the cell size (grain) of the study.

However, the choice of cell size may also be determined by other criteria. For instance, a larger cell size might result in a more manageable data set or might be chosen if spatial autocorrelation is measured within the species' data and, as a result, observations that cannot be considered independent need to be aggregated. In contrast, a smaller cell size might better represent the ecological processes. Here, we will focus mainly on situations where spatial accuracy is known and can vary from one observation to the other.

Data of varying spatial accuracy can be manipulated to avoid propagating measurement errors in the model (Elith, Burgman & Regan 2002) by either (i) aggregating all data in regular grid cells (or possibly other cell shapes) whose size matches the poorest locational accuracy

of observed occurrences, or (ii) dropping the most inaccurate data. A balance offering the best sample size vs. accuracy is usually found between these two options.

This is illustrated in Fig. 1, which shows the decrease in the number of occurrences in Switzerland of the rare species *Eryngium alpinum* L. (*Apiaceae*) as the spatial resolution increases (i.e. decreasing cell size). This is due to the fact that fewer occurrences have a high locational accuracy associated with grid cells at high resolution (fine grain). Lowering the spatial resolution (coarser grain) allows less precise observations to be made, thus increasing their overall number. However, such a decrease is not necessarily straightforward because, when lowering the resolution (i.e. increasing cell size), distinct occurrences can also become aggregated in the same grid cell. Hence, the choice of method depends on the resolution of environmental layers available in the GIS, on the biology of the focus species and on the spatial distribution of its recorded occurrences.

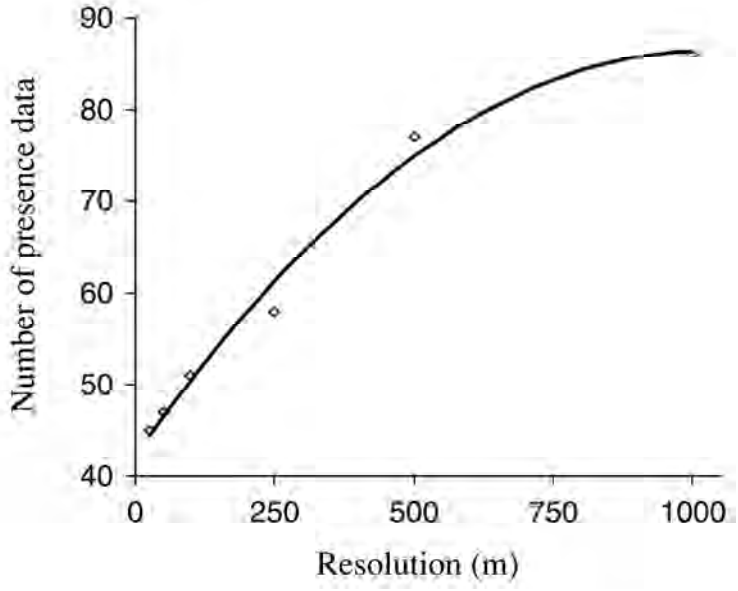


Fig. 1. Number of occurrences of *Eryngium alpinum* in Switzerland at each resolution (cell size).

Another common problem with data that is retrieved from existing databases is that it often lacks absence observations. Such presence-only data severely limits the fitting of many statistical models. For instance, generalized linear models require both presences and absences for their calibration (Guisan, Edwards & Hastie 2002). One possibility to get around the lack of absence data is to use models based on presence-only data. These are called profile techniques, as opposed to group discrimination approaches that need presence-absence or abundance data (Robertson, Caithness & Villet 2001). A well-known example of a profile-type model is the climatic envelope approach developed largely in the late 1980s by Australian scientists and implemented in the BIOCLIM package (Busby 1991; now ANUCLIM; Houlder et al. 2000). Another, more recent, example is the ecological niche factor analysis (ENFA) implemented in the BIOMAPPER package (Hirzel et al. 2002a; Hirzel, Hausser & Perrin 2002b). However, a common problem of profile methods is that they tend to generate overoptimistic predictions, i.e. they predict the species at too many locations. This is easily understood by the fact that, from a quantitative evaluation perspective, a 'perfect' model with such data would be a model predicting the species everywhere (i.e. '1' would be attributed to all cells in the area), as all observations would be correctly predicted as '1' and no discriminating absence would be available to restrict the predictions to zero where needed (i.e. at environmentally inappropriate locations).

In this regard, GLM constitute a better choice because they can deal with many types of predictors (continuous, binary, qualitative, ordinal), but on the other hand they must have presence and absence data. In order to use GLM when no absence data are available, one approach is to generate 'pseudo-absences' (Zaniewski, Lehmann & Overton 2002) and to use them in the model as absence data for the species. The manner in which pseudo-absences

are generated is particularly important because it can have a significant influence on the final quality of the model (Zaniewski, Lehmann & Overton 2002).

The easiest way to choose pseudo-absences is simply to generate them totally at random over the study area (Hirzel, Helfer & Métral 2001; Zaniewski, Lehmann & Overton 2002). However, this method runs the risk of generating an absence in an area that is, in fact, favourable to the species. Indeed, when dealing with common species, choosing such a 'wrong absence' may not be too problematical because the numerous presence records will counteract its effect. However, when working with rare species, data are often scarce and choosing a wrong absence could significantly reduce the quality of a model.

To avoid, or at least reduce, this problem, more subtle methods can be employed to generate the pseudo-absences. For example, Zaniewski et al. (2002) first create a habitat suitability (HS) map of all fern species they wish to model (i.e., pooling together presence data of all their species) using a GAM with totally random pseudo-absences. Then, they randomly select a second set of pseudo-absences proportionally to the predictions of their first HS map and use it to fit GAMs for every species individually. Selecting pseudo-absences proportionally to the overall sampling effort aims at avoiding sampling pseudo-absences in sites that were under-sampled in the field. Unfortunately, multi-species data are not always available. In such situation, the first map – based on purely random pseudo-absences – is specific to the modelled species and pseudo-absences can be selected in areas below a certain threshold of habitat suitability, in order to maximize the discriminating ability of the second model. The choice of this threshold must be defined as objectively as possible, for instance as the lowest value still encompassing 95% of observed species' occurrences.

In this study, we propose another way to generate pseudo-absences, which combines the respective strengths of ENFA and GLM. It is also a two-step approach, but uses ENFA instead of a GLM with totally random pseudo-absences to calculate the first HS map that is used to weight the selection of pseudo-absences. The calculation of this first model is particularly straightforward with ENFA (e.g. no need to select predictors).

The aims of this study were twofold. The first objective was to evaluate different methods for predicting rare species distribution, using ENFA with presence-only data, GLM with presence and random pseudo-absences, and a combination of both approaches. The second aim was to assess the dilemma between data quality and quantity, trying more specifically to answer the question: is it preferable to have a large number of observations, which is better from a statistical point of view, or should one favour locational accuracy of observations (dropping all inaccurate ones, thus using a reduced set to calibrate the model) to ensure a better correspondence with environmental predictors used to predict the observations? This part of the study was conducted by building models at two different resolutions (25 and 500 m) having a different number of occurrences associated with each (Fig. 1). *Eryngium alpinum* (*Apiaceae*), a flagship threatened species in the European Alps, was chosen as an illustration. Finally, results from field investigations demonstrate the usefulness of such a model for suggesting new observation sites for rare and endangered species.

Methods

Study Area

Switzerland covers an area of 41'293 km2 and consists essentially of two mountain chains with a west–east orientation: the Jura (highest peak in Switzerland 1607 m a.s.l.) in the north and the Alps (highest peak in Switzerland 4634 m a.s.l.) in the south, separated by a lowland corridor, 50–100 km wide, generally referred to as the Swiss midlands and ranging from about 360 to 900 m a.s.l.

The whole country belongs to a single floristic unit, the medio-european or subatlantic domain (Ozenda 1982), which also corresponds to a single general climate of suboceanic type. A large proportion of the rain carried from the Atlantic Ocean and North Sea is stopped by the mountains and thus highest annual rainfall occurs in the Jura and in the northern part of the Alps. Mean annual temperature is also unequally distributed, with warm, wet areas in the south-eastern part of the country (Ticino, southern Alps), warm, dry areas in the south-eastern part (Valais), and a cooler and wetter climate characterizing the midlands and the Jura.

Test Species

Eryngium alpinum L., commonly called Alpine Eryngo or Queen of the Alps, is a perennial hemi-cryptophyte about 30–100 cm high and easily recognizable by its cylindrical capitulum surrounded by a large involucre of huge blue-violet bracts. According to Landolt (1977), *E. alpinum* prefers habitat with moist, deep, alkaline soils rich in nutrients. It also needs full sunlight and therefore almost never occurs in forested areas.

Eryngium alpinum often grows on steep slopes of the alpine and subalpine belts, in a megaphorb-like vegetation. In Switzerland, it is found between 1400 and 2100 m, essentially in the pre-Alps (the northern foothills of the Alps) of the cantons of Vaud, Valais and Fribourg but some populations are also found in the cantons of Grisons, Uri and Unterwald (eastern part of Switzerland). Due to its ornamental features, this species is also often grown in gardens and cemeteries. Although the exact reasons for the recent decline of this species are still unknown, major threats are thought to be picking, and changes in pasturing practices. The species is threatened in the whole of Switzerland and appears on the Swiss red lists of endangered plant species (Moser et al. 2002). The genetic structure and reproductive biology of *E. alpinum* are the focus of research to understand better the causes of its rarity (Gaudeul et al. 2002).

Species Data

The Swiss Floristic Network (CRSF) in Geneva provided a large part of the data used in this study. These data consist of observations of presences only (hereafter referred to as presence-only data) of *E. alpinum* at known geographical locations. These observations were aggregated within regular square units of two grids covering the country, resulting in 46 and 77 occurrences at resolutions of respectively 25 and 500 m (Fig. 1). No geographical location of observed absence was available at the start of the study.

Environmental Data

Hereafter, environmental variables used to predict species distribution are termed the predictors. On a mesoscale such as the whole of Switzerland, direct environmental predictors such as climate should prove more powerful than indirect predictors like altitude (Guisan & Hofer 2003). This is because different climates, not all of which might be suitable for the target species, can occur at the same altitude throughout Switzerland. Due to the need for normality in the ENFA method, only quantitative predictors were used. As a comparison, the same predictors were used to fit the GLM, although the latter statistical method can theoretically deal with all kind of predictors.

The following pool of quantitative environmental predictors was selected according to ecology and data availability for *E. alpinum* in the ArcGIS software (ESRI Inc., Redlands, CA, USA) in the case of GLM-based predictions and in the IDRISI software (Eastman 1997) in the case of ENFA-based predictions (Table 1).

The term pool is used here to indicate that not all of these environmental predictors were necessarily used to fit the different models. The original resolution of all environmental maps was of 25 m. That is, over the map of Switzerland, the data layers were divided into pixels of 25 × 25 m. To predict the species distribution at the wider resolution of 500 × 500 m, an aggregation of the 25 × 25-m data was performed in ArcGIS by calculating the average value of the 400 25-m pixels enclosed within each 500 × 500-m pixel.

In addition, two qualitative environmental variables providing information, respectively, on main land-use classes (land-use) and on geology and soil types (substratum) associated with each pixel were used separately in a last step, to filter predictions made by the models.

Table 1. Descriptions and abbreviations of the quantitative environmental variables forming the initial pool of predictors. They were all originally prepared at a resolution of 25 m. The last two rows represent two qualitative variables. A few selected classes of these were used to filter the final predictions (see text).

Abbreviation	Description	Unit
slope	Average slope of each quadrate.	%
Srad3	Sum of March solar radiation.	$kJ \times day^{-1}$
Srad7	Sum of July solar radiation.	$kJ \times day^{-1}$
Tave7	Average July temperature.	°C
Rain48	Sum of rainfall from April to August.	mm
Rain49	Sum of rainfall from April to September.	mm
ddeg300	Sum of daily average temperature above 3°C.	°C
Topo500	Topographical position with a radius of 500 m.	m
Topo1000	Topographical position with a radius of 1 km.	m
Landuse	Land-use classes (forests, agricultural areas, roads, buildings, rivers, open areas, screes, glaciers, etc.), rasterized from a 1:25 000 topographic vector map (Vector 25).	43 classes
Substratum	Geotechnical map providing information on geology and soil, rasterized from a 1:200 000 vector map (Geotech).	30 classes

Statistical Methods

The two statistical methods used within this study were the ENFA (Ecological Niche Factor Analysis) and GLM (Generalized Linear Model).

Ecological niche factor analysis (ENFA):

The ENFA (Hirzel et al. 2002a) is a method based on a comparison between the environmental niche of the species and the environmental characteristics of the entire study area (stored as GIS layers), hereafter termed the background data. Hence, ENFA only needs a set of presence data (no absences are required) and a set of background GIS predictors. In contrast to many other predictive methods that can be fitted outside the GIS (like GLM), ENFA thus requires a dynamic access to the ecogeographical predictors.

ENFA is similar to a principal component analysis (PCA) in that it also transforms the original ecogeographical variables into new, uncorrelated, axes. However, whereas in PCA the successive axes are simply selected to match the direction of maximum variance in the multidimensional ecogeographical space, ENFA's principal components all possess a true ecological meaning for the modelled species.

The first component is called the marginality factor (MF). It passes through the centroid of all species' observations (multidimensional optimum) and the centroid of all background cells in the study area (mean environmental conditions). Hence, a high marginality value indicates that the species' requirements differ considerably from the average habitat conditions in the study area.

Several specialization factors (SF) are then successively extracted from the n – 1 residual dimensions. Each SF is calculated in order to (i) ensure its orthogonality to the marginality factor and to the other SF, while at the same time (ii) maximize the ratio between the residual variance of the background data and the variance of the species' occurrences. A high specialization indicates that a species has a restricted ecological tolerance compared with the overall range of conditions that prevail in the study area.

Because most of the information is usually contained in a few first factors (usually the marginality factor and up to three or four specialization factors), only these are kept to compute the final HS map. All cells in the map obtain a HS value that is proportional to the distance between their position and the position of the species optimum in the new factorial space.

All ENFA analyses were performed within the BIOMAPPER software (version 2.1; Hirzel, Hausser & Perrin 2002b). Correlations between all variables in the initial pool of predictors (Table 1) were calculated prior to ENFA analyses, in order to determine which variables should preferably be used in the ENFA. When two or more predictors had a correlation coefficient greater than 0.5, only the most proximal (in the sense of Austin 2002) was kept for the ENFA. As ENFA requires normally distributed data, all environmental layers were normalized through the 'box–cox' algorithm (Sokal & Rohlf 1981). Although several variables did not recover normality after the box-cox normalization, ENFA is not considered too sensitive to such violation (Hirzel et al. 2002a).

Generalized Linear Models (GLM):

GLMs (McCullagh & Nelder 1989; for HS application of GLM see Guisan, Edwards & Hastie 2002) are an extension of the classical multiple regression, allowing non-normal response variables to be modelled. GLM were used in our case to model presence-absence of the species. As absences were not available in the original data set, pseudo-absences were generated in various ways (see below). All GLM were fitted within the R software (R 1.4.0; A Language for Data Analysis and Graphics ©2002), by specifying a binomial distribution and a logistic link function, as similarly done for other presence–absence data in ecological studies (Guisan, Weiss & Weiss 1999; Manel, Dias & Ormerod 1999; Guisan & Hofer 2003; also called logit regression).

Selection of predictors (and their possible transformations, e.g. polynomial terms) is certainly the most important and difficult step when fitting a GLM. As the number of combinations is too great to test all of them, we used a custom stepwise selection procedure programmed in R that offers the best-suited combination of predictors, even with large data sets. A first exploratory analysis of the different predictors based on univariate GLM (i.e. fitting a single predictor at a time but allowing polynomial terms to be considered) showed that all response curves were at most of a $y = x + x^2$ type, providing a basis for ignoring polynomial terms higher than quadratic. In a second step, GLM were fitted to all possible pairs of uncorrelated predictors (and their square term if significant) and the reduction of deviance was tested and recorded in each case. The pair of predictors causing the highest deviance reduction was kept. In the third step, deviance reduction was tested on each pair of predictors previously selected, adding the remaining predictors one at a time. Again, the trio expressing the greatest deviance was kept. Finally, step three was repeated until the addition of any predictor, and possibly its square term, was no longer significant. This procedure is close to forward stepwise selection, although a significant difference is that the pair of predictors (among all possible pairs) causing the highest deviance reduction is entered first in the model, and the same rule further applies to the selection of the following predictors in the formula.

The best combination of predictors was considered to be the one that explained the greatest amount of deviance while having all individual terms in the equation explaining each a significant amount of deviance ($p \leq 0.05$, chi-test in the case of binomial models).

The Modelling Framework

Three methods were used to model species' distributions from presence-only data (i.e. without observed absences): (i) ENFA; (ii) GLM with randomly chosen pseudo-absences, hereafter called GLM-R; and (iii) GLM with ENFA-weighted pseudo-absences, hereafter called GLM-ENFA (Fig. 2). To assess the importance of quantity vs. quality, these models were all fitted with data prepared at the two resolutions of 25 m and 500 m.

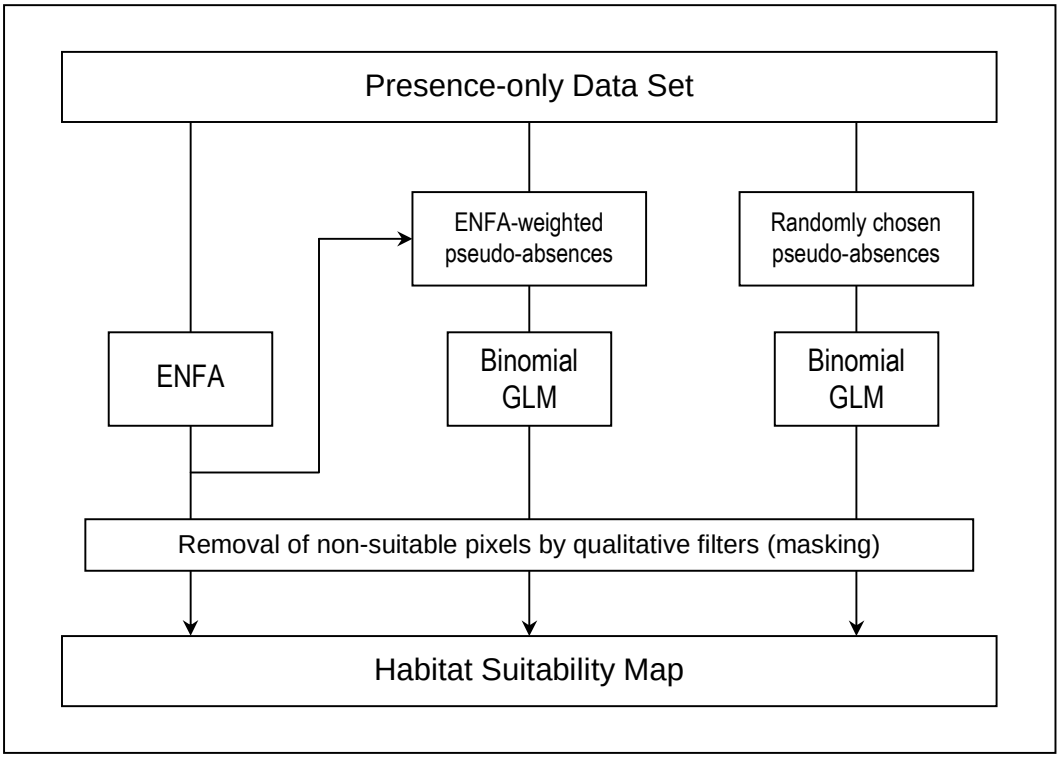


Fig. 2. The three methods used to produce habitat suitability maps: ENFA, GLM with totally random selected pseudo-absences, and the combined GLM-ENFA method where the pseudo-absences are weighted by the result of the ENFA.

The generation of pseudo-absences was done in two ways. (i) Totally at random: geographic coordinates were chosen at random over the Swiss territory. (ii) Weighted by ENFA predictions: in this case, coordinates were also chosen at random, but only in areas where predictions by the prior ENFA model for *E. alpinum* were lower than 0.3. This threshold was chosen because it was slightly lower than the lowest ENFA prediction of 0.33 associated with observed presences.

In each case, the number of generated pseudo-absences was the same as the number of real presences, namely 45 for the 25-m resolution and 77 for the 500-m resolution (Fig. 1). In addition, pseudo-absences were never chosen in pixels having a value for land-use or substratum that was incompatible with the presence of *E. alpinum*, because these areas were removed later on during the filtering procedure. Moreover, as particular environmental factor values can sometimes be associated with these incompatible areas (e.g. glaciers usually have lower temperatures than surrounding areas) this might reduce the discriminating power of the model (i.e. the model will differentiate between suitable and very unsuitable

locations but will not highlight finer variants in more or less suitable areas). Pseudo-absences were then combined with real presences into a single presence-absence data set ready to use in a binomial GLM.

Because chance plays a part in the choice of the pseudo-absences, each modelling procedure was repeated 1500 times (for both types of pseudo-absence and for both resolutions, 25 and 500 m). For each of these procedures ($4 \times 1500 = 6000$ in total), the custom stepwise selection was performed, ensuring that the best predictor combination was selected for each data set (as they all differed in their pseudo-absences).

Finally, qualitative environmental data (i.e. landuse and substratum layers) were used to filter the ENFA-based and GLM-based predictions in order to eliminate those areas where *E. alpinum* is unlikely to grow, e.g. urban areas, forests or glaciers. Such end-process filtering was performed by setting the model predictions to zero where unsuitable land-use and substratum classes occurred.

Model Evaluation

The following evaluation measures were calculated for each of the 1500 GLM fitted on each presence/pseudo-absence data set, at each resolution.

Explained deviance (*adjusted-D^2*): This is the percentage of deviance (i.e. variance in GLM) explained by the GLM. This measure expresses the fit of the model, weighted by the effective number of degrees of freedom (i.e. taking into account the number of predictors and the number of observations) used to build the model (Guisan, Weiss & Weiss 1999).

Best kappa (*B-kappa*): The kappa coefficient (Cohen 1960; Fielding & Bell 1997) was calculated for all thresholds between zero and one by increments of 0.05. The greatest value was kept as the 'best kappa' value (Elith 2002). This measure expresses the best possible agreement not obtained randomly between two qualitative variables (of which a binary variable is a particular case).

Gini coefficient (Eq. 1): This is a transformation of the area under the curve value (AUC), obtained from a receiver-operating characteristic plot method (ROC-plot; Fielding & Bell 1997), so that values have a wider range than the AUC (Hand & Henley 1997; Copas 1999):

$$Gini\ coefficient = 2 \times (AUC - 0.5) \qquad \text{(Eq. 1)}$$

The Gini coefficient as used here is an extension of its original use to describe income disparity. It will usually vary between zero (for an uninformative model) and one (for a model with perfect discrimination), but exceptionally it could be negative for cases where the model tends to make higher predictions at absence sites (i.e. the model is worse than chance; J. Elith, personal communication).

Minimal predicted area (MPA): The minimal predicted area (MPA) is the minimal surface obtained by considering all pixels with predictions above a defined probability threshold (e.g. 0.7) that still encompasses 90% of the species' occurrences (see more explanation below).

To our knowledge, the MPA measure has never been used before and requires more explanation. When evaluating HS maps using presence-only data, a map predicting presence of the species everywhere would show the best evaluation (because all presences would then effectively be predicted as presences), but such a map would be useless. The idea behind MPA is based on the assumption that a good HS map drawn from presence-only data

should predict a potential area that is as small as possible (rule of parsimony) while still including a maximum number of the species' occurrences. For each model, predicted values were attributed to all occurrence sites and the value (e.g. 0.7) above which 90% of occurrences were observed was used as a threshold to reclassify the probability HS map into presence/absence (1/0). The positively predicted area (i.e. where the model's projections are above the threshold encompassing 90% of observed occurrences) corresponds to the 'minimal predicted area'. All MPA values were calculated before the filtering operation because we wanted to assess performances of the statistical models only.

As no independent presence–absence data set was available, a common situation with rare species, best kappa and Gini coefficient were calculated on pseudo-independent presence/pseudo-absence data sets generated through a 'leave-one-out' jack-knife procedure (Jaberg & Guisan 2001). These two measures could not be calculated in the case of simple ENFA models, because no pseudo-absences were necessary for building these models. This is a recurrent problem encountered when evaluating model predictions with presence-only data. As a result, MPA was the only evaluation measure available for ENFA models.

The evaluation of GLM-based methods (i.e. GLM-ENFA and GLM-R) was performed on each of the 1500 generated data at both resolutions using the four evaluation measures, whereas the evaluation of the ENFA was performed on the single HS map calculated at each resolution. Hence, evaluation values of GLM-ENFA and GLM-R approaches are best provided in the form of box-plots while ENFA evaluations are represented by single values (bars) in the MPA graphs (Fig. 3).

Comparing the different evaluation measures

We further tested whether those models (out of the 1500 runs for each of the four GLM experiments) that obtained the best evaluation for explained deviance, best kappa and Gini coefficient, were also those that obtained the best MPA evaluation. To test this, correlation coefficients (Spearman) were calculated between the ranks given to each model by the different evaluations for both GLM-ENFA and GLM-R at both resolutions.

Using HS maps to support field investigations

The final goal of our study was to predict and discover new populations of *E. alpinum*. For this purpose, we selected the GLM-ENFA model that obtained the best average value for explained deviance, the best kappa and Gini coefficient (i.e. a rank classification was made for each of these evaluation measurements and the model with the best average rank was chosen). This model was implemented in the ArcGIS software and the resulting probability map was filtered with the discriminating classes of qualitative predictors (filters) to remove areas where the presence of *E. alpinum* was highly improbable (i.e. with no occurrence ever recorded).

The predictions of the models that fell within the following types of land-use classes were set to zero (i.e. unsuitable habitat): forests, urban and agricultural areas, lakes and rivers, glaciers, swamps and other areas transformed by humans (i.e. gravel pits, dams, etc.). The same operation was performed using the substratum map, and only seven classes of limestone and marly substratum were considered suitable for the species, all others being set to zero.

Results

Correlations between environmental predictors calculated at the 25-m resolution were very similar to those calculated at the 500-m resolution, so that the predictors retained for ENFA analyses at both resolutions were the same: slope, srad3, ddeg300, rain49 and topo500 (for their descriptions see Table 1).

The HS map was obtained by considering the first two components of the ENFA, which expressed, respectively, 92.8% and 88.1% of the variance at the resolutions of 25 m and 500 m. MPA values obtained for the two ENFA HS maps are given in Fig. 3d. For the two types of GLMs, box-plots in Fig. 3 show the distribution of (a) explained deviance (D^2), (b) best kappa value (*B-kappa*), (c) Gini coefficient (*Gini*) and (d) MPA.

A Wilcoxon rank test confirmed that, for all these evaluation values, the averages were significantly different ($P < 0.01$) between the following pairs of models: GLM-ENFA 25 : GLM-R 25, GLM-ENFA 500 : GLM-R 500, GLM-ENFA 25 : GLM-ENFA 500 and GLM-R 25 : GLM-R 500. The number indicates the spatial resolution considered. Correlations between the different measures of evaluation calculated for each fitted model at each resolution are given in Table 2.

Table 2. Spearman rank correlation coefficients between the different evaluation values calculated for each fitted model. The correlation values are averages of the correlations obtained for GLM-ENFA and GLM-R methods at both 25-m and 500-m resolutions (n = 2 × 2 × 1500 = 6000 models)

	Explained deviance	Best kappa value	Gini coefficient	MPA
Explained deviance		0.76	0.87	-0.01
Best kappa value			0.74	-0.21
Gini coefficient				-0.12
MPA				

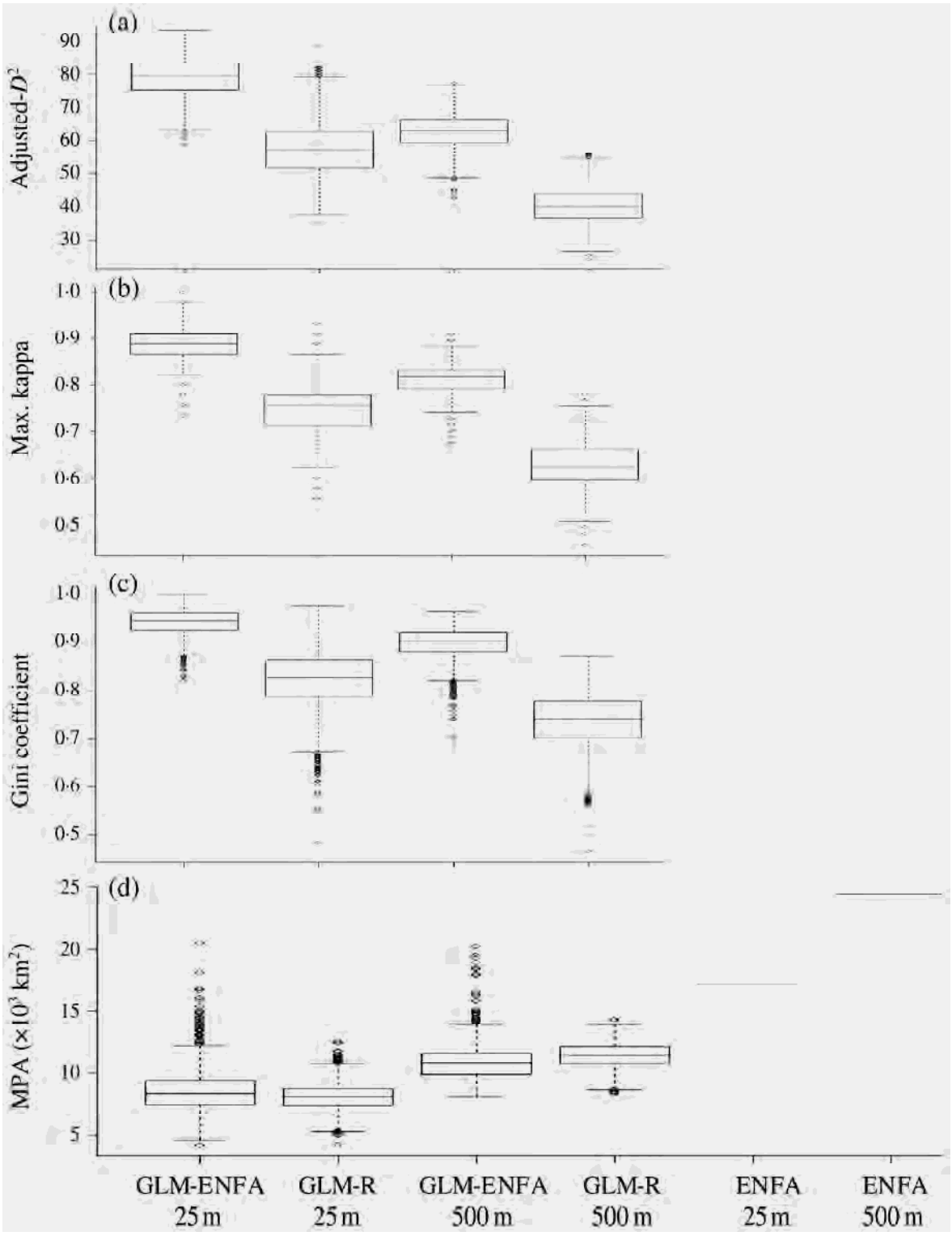


Fig. 3. Box-plots of the four evaluation measures calculated for each method of creating HS maps: **(a)** explained deviance in percent (adjusted for the effective number of degrees of freedom); **(b)** maximum kappa value; **(c)** Gini coefficient; **(d)** MPA value (in 10^3 km^2). Values for ENFA are only shown for MPA as other values could not be calculated because of the lack of real absences. On graphs (a-c), the higher the value of evaluation, the better the model. On graph (d), the lower the MPA the better, but in conjunction with the highest possible value for the evaluation measures (a-c).

Based on the cartographic implementation (potential map; Fig. 4) of the GLM-ENFA model at a resolution of 25 m, four new populations of the species were discovered in the field, all of them in pixels characterized by a high to very high habitat suitability (probability values of 0.98, 0.93, 0.92 and 0.79).

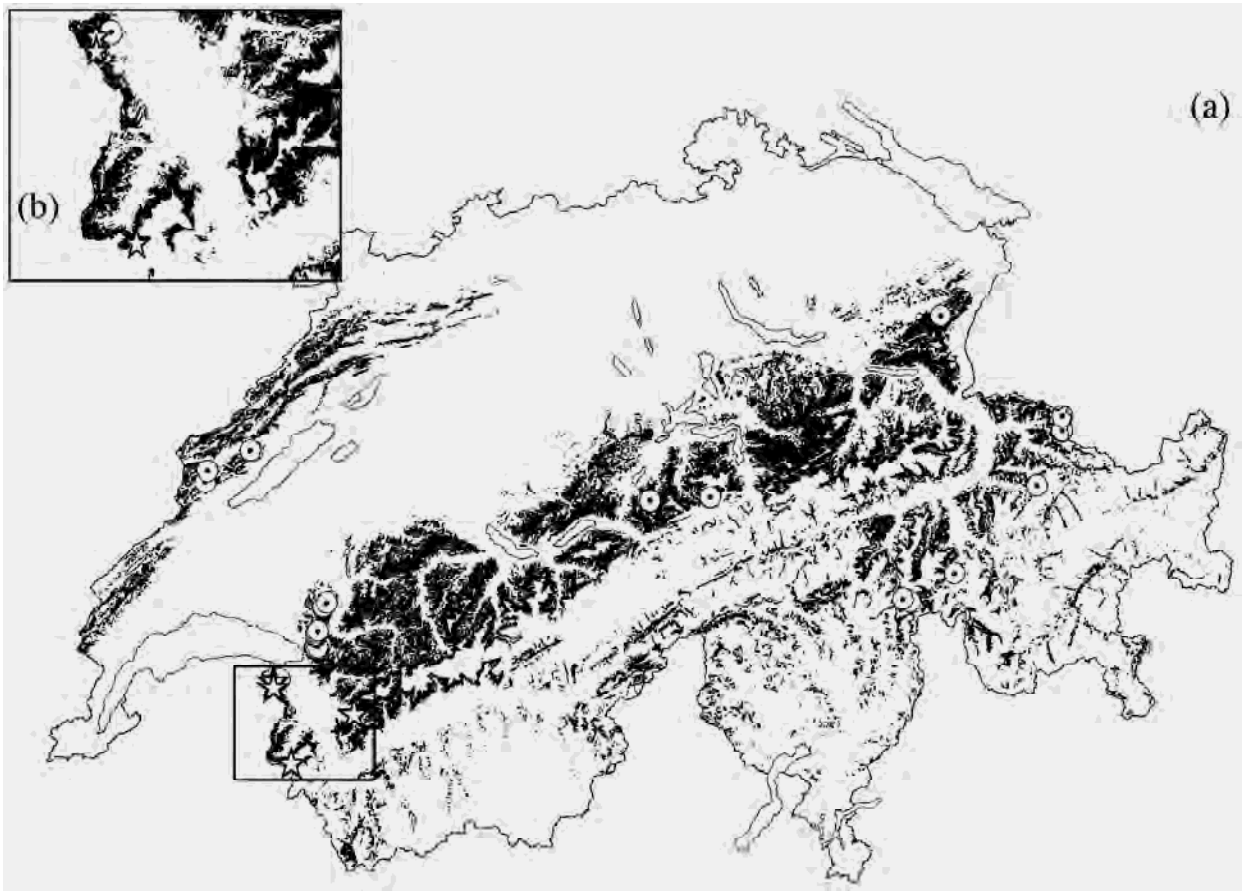


Fig. 4. (a) Potential distribution map for *Eryngium alpinum* in Switzerland drawn from one of the GLM-ENFA models at a resolution of 25 m. Black and dark grey tones indicate areas projected as highly suitable. White circles indicate real presence points used to generate the map and white stars represent two new populations of the species discovered in the field. Highly suitable areas tend to be located in mountainous regions with higher rainfall (Jura and northern part of the Alps), which is consistent with the ecology of *E. alpinum*. **(b)** Magnified map showing the locations (star symbols) of the four new populations. Note that these are located within areas projected as highly suitable (see text).

Discussion

Our first goal was to compare two existing and one new approach to predicting rare species distribution with presence-only (occurrence) data. Due to the lack of true absences, a formal comparison between ENFA and GLM-based methods (i.e. GLM-ENFA and GLM-R) is difficult. Indeed, in our study, MPA is the only evaluation measure available for comparison. Results show that the HS maps obtained with ENFA predict a MPA value that is approximately twice the mean of the GLM-based methods, both at the 25-m and 500-m resolutions (Fig. 3d). This result seems to confirm the tendency of ENFA to over-predict species distribution (Zaniewski, Lehmann & Overton 2002), due to the lack of discriminating absences, as discussed in the introduction. Another possible explanation of the apparent (but not proved) lack of accuracy of ENFA models could result from a violation of the assumption of normality of predictors that is required by the ENFA algorithm, as many of our predictors were actually not normally distributed (Kolmogorov–Smirnov goodness-of-fit test). Further testing would be needed to assess the robustness of ENFA in such circumstances (Hirzel, Helfer & Métral 2001, Hirzel et al. 2002a). This situation is likely to prevail in many other similar studies.

However, although we observed a large difference in MPA values between ENFA and GLM-based methods, it should not be concluded that the latter methods always prove better than the former. For instance, Hirzel, Helfer & Métral (2001), using a virtual species with known absences in a real landscape, have shown that ENFA can prove superior to GLM in the specific case of invading species that do not yet occupy all their potential habitats in the landscape. Such non-equilibrium between species distribution and environment might be less likely in the case of rare and threatened species, which tend to occupy most of their potential habitats, as these have usually been drastically reduced and, as a result, cover only a small proportion of the territory. However, although *E. alpinum* is truly a rare species (in the sense of Rabinovitz et al. 1986), it always yielded rather spatially large projections of suitable habitat (compared with other species; O. Broennimann & A. Guisan, unpublished results), which suggests that either its habitat is not so spatially restricted and that other reasons (like cutting) are limiting its distribution, and/or that important environmental variables are missing in our model.

More detailed comparison was possible between GLM-ENFA and GLM-R methods because they are both based on logistic regression and thus use presence and absence data. However, care should be taken when interpreting these results, as our evaluation measures are based on pseudo-absences and not real observed absences. This is a recurrent limitation of evaluating models based on occurrence data and a research area where progress is still required.

The three first evaluation measures (*Adjusted-D^2*, *best-kappa* and *Gini coefficient*, Fig. 3a-c) are consistent with each other, showing that GLM using ENFA-weighted pseudo-absences provide significantly better results on average (Wilcoxon rank test) and less variability than those using randomly chosen pseudo-absences. This is true at both the 25-m and 500-m resolutions, which confirms that choosing pseudo-absences in an ENFA-weighted way rather than totally at random will enhance the accuracy of an HS map.

Interestingly, results from MPA measures were not consistent with the other evaluation measures. Indeed, we did not expect such consistency because the MPA concept is based on the parsimony criterion that 'the smaller the potential map the better the related model', which does not necessarily fit the mathematical criterion of statistical evaluations. At the higher resolution (25 m) GLM-R models provided a smaller MPA average value (remember that for MPA, the smaller the value the better the model) and a smaller deviance, whereas at

the lower resolution (500 m) GLM-ENFA models obtained a similar average MPA value as GLM-R models, but showed a much larger deviance.

Furthermore, the comparison of the four evaluation measures, based on all individual model predictions (1500) in each GLM experiment (two types of GLM at two resolutions), did not show any correlation between MPA and any of the three quantitative measures (D^2, *best-kappa* or *Gini coefficient*). We do not believe that these results imply that MPA is a non-reliable value because the rule of parsimony used in MPA has a practical justification in conservation ecology, especially in the case of rare species that are, by definition, not widely distributed. Hence, further research is needed, at least in the case of these species, to assess whether the best HS map would not be the one that maximizes quantitative evaluation statistics while at the same time minimizing the predicted area. One identified limitation to the generalized use of MPA might be that it fails to evaluate appropriately those potential maps produced by certain modelling techniques, such as BIOCLIM (Busby 1991), as this type of model always fits the maximum possible predictions at observed presence sites (J. Elith, personal communication).

The second goal of this study was to determine whether it is preferable to use (i) less data with higher spatial accuracy or (ii) more data with lower accuracy. Comparing the evaluations of the 25-m and 500-m resolution HS maps reveals that averages for all these evaluation values are always better with the 25-m resolution. Overall, this conclusion is still valid when considering the three different types of models, GLM-ENFA, GLM-R and ENFA, although their deviances do not differ significantly.

The lower performances observed at the 500-m resolution could result from the combination of three factors. First, a loss of information is inevitable when aggregating environmental variables. Secondly, the low accuracy associated with some species' occurrences used at this resolution might still be underestimated and a greater measurement error (as defined by Elith, Burgman & Regan 2002) might actually prevail in the data. This is less likely to occur in the case of observations that have a higher spatial accuracy. Thirdly, plants being fixed organisms, they are highly influenced by the local microclimate. Therefore, relating species data that have a high geographical precision (of site location) with high-resolution environmental data should have a better predictive power because they reflect very local ecological conditions, while aggregated data reflect smoother environmental gradients in the area (e.g. mesoclimate). Furthermore, some important combinations of environmental predictors might not be appropriately expressed in such aggregated data.

In turn, such superiority of higher resolution predictors and less data might not be true for non-fixed organisms, as the required resolution for these is certainly dependent on a larger home range of target species, as suggested for bats by Jaberg & Guisan (2001). None the less, many of our potential maps have spatial predictions that cover a large proportion of the mountainous areas of Switzerland, even those with a good evaluation. This primarily reflects the fact that large areas are probably truly suitable sites for the target species *E. alpinum*, from the single perspective of those predictors that were used to build the model. Other factors, not included in the model but which might potentially have a more direct influence on plants (i.e. more proximal in the sense of Austin 2002), probably account for the unexplained spatial variation, and thus could enhance model accuracy. The problem remains, however, that data on many of these very important environmental factors, such as nutrient content in soils or precise physiologically meaningful microclimatic measurements, are still difficult to obtain in a spatially explicit way.

The best option for improving the HS maps would certainly be to obtain additional data for the target species, but this is difficult in the case of rare species. HS maps prove

particularly useful in this regard, by suggesting new sampling sites for the species, such as those pixels of high prediction values that are not in the close vicinity of an observed population of the species. Visiting such suggested sites in the field at the optimum flowering time for the species may produce new presences or, at least, attested absences. This was done at the end of our study and four new populations of *E. alpinum* were found at sites of high predicted probability of presence. Such new data should then optimally be used to refine the models and generate new predictions that will need to be verified in the field. Theoretically, reiterating such processes should lead to equilibrium, when new data (presences or absences) no longer contribute to improvement of the model (reaching a plateau). Another solution for improving the accuracy of HS maps could be to refine the choice of pseudo-absences. In this study, the GLM-ENFA method was indeed used in a relatively simple manner. We used the ENFA predictions to divide the study area into two parts, one with probabilities of presence greater than 30% and the other with lower probabilities. Pseudo-absences were then randomly chosen in the latter category. A more subtle way of choosing the pseudo-absences could be, for instance, to stratify their distribution along a suitability gradient, like mean annual temperature. This could be a more precise method for choosing suitable areas.

An alternative to the process of selecting as many pseudo-absences as there are presences (usually a very limited number for rare species), and repeating the process a large number of times (e.g. 1500, as done here), might be to sample, once only, a larger number of pseudo-absences (say 10'000) and assign them a weight ($w = k/10'000$) in the GLM so that the sum of the weight of all pseudo-absences adds up to the sum of presences k, thereby ensuring a prevalence of 0.5 for presences and absences respectively (M. Wisz, personal communication). This could prevent the inherent risk of inappropriately choosing a limited number of random pseudo-absences (i.e. leading to a poor model fit) when only one random selection run is made.

Our third goal was to use the obtained potential distribution maps for suggesting new sampling sites for rare species. Although this study was not focused on this particular application of predictive models, a preliminary field campaign based on the selected GLM-ENFA potential distribution map (Fig. 4) led to the discovery of four new populations of this highly endangered species. Indeed, this is a very promising result that strongly supports the use of predictive habitat distribution models for nature conservation planning.

Acknowledgements

To Jane Elith, Alan Fielding, Steve Rushton, Mary Wisz, Anthony Lehmann and an anonymous referee for their useful comments. To Simone Peverelli and Christophe Randin for their technical help with GIS. To Beat Bäumler at the Center of the Swiss Floristic Network (CRSF) in Geneva for providing most species data. We are also very thankful to Niklaus E. Zimmermann and Felix Kienast for providing the climatic data and Yamama Naciri-Graven, Olivier Broennimann, Mathias Vust and Charlotte Trippi for providing additional species data. Finally, to Julie Warrillow and Sophie Rickebush for kindly making the linguistic revision.

References

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101-118.

Bakkenes, M., Alkemade, J.R.M., 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.

Brown, D.G. (1994) Predicting vegetation types at treeline using topography and biophysical disturbance variables. *Journal of Vegetation Science*, **5**, 641-656.

Brzeziecki, B., Kienast, F. & Wildi, O. (1993) A simulated map of the potential natural forest vegetation of Switzerland. *Journal of Vegetation Science*, **4**, 499-508.

Busby, J.R. (1991) BIOCLIM - a bioclimate analysis and prediction system. Nature Conservation: Cost Effective Biological Surveys and Data Analysis (eds C.R. Margules & M.P. Austin), pp. 64-68. CSIRO, Melbourne, Australia.

Carey, P.D. & Brown, N.J. (1995) The use of GIS to identify sites that will become suitable for a rare orchid, Himantoglossum hircinum L., in future changed climate. *Biodiversity Letters*, **2**, 117-123.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **41**, 687-699.

Copas, J. (1999) The effectiveness of risk scores: the logit rank plot. *Journal of the Royal Statistical Society Series C-Applied Statistics*, **48**, 165-183.

Eastman, J.R. (1997) Idrisi for Windows. Clark University, Worcester, UK.

Elith, J. (2002) Predicting the distribution of plants. PhD Thesis. University of Melbourne, Melbourne, Australia.

Elith, J. & Burgman, M. (2002) Predictions and their validation: rare plants in the central highlands, Victoria, Australia. Predicting Species Occurrences: Issues of Accuracy and Scale (eds J.M. Scott, P.J. Heglund, J.B. Haufler, M. Morrison, M.G. Raphael, W.B. Wall & F. Samson), pp. 303-313. Island Press, Covelo, CA.

Elith, J., Burgman, M.A. & Regan, H.M. (2002) Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling*, **157**, 313-330.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38-49.

Franklin, J. (1995) Predictive vegetation mapping: geographical modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, 474-499.

Gaudeul, M., Naciri-Graven, Y., Gauthier, P. & Pompanon, F. (2002) Isolation and characterization of microsatellite markers in a perennial Apiaceae, *Eryngium alpinum* L. *Molecular Ecology Notes*, **2**, 107-109.

Godown, M.E. & Peterson, A.T. (2000) Preliminary distributional analysis of US endangered bird species. *Biodiversity and Conservation*, **9**, 1313-1322.

Guisan, A. & Hofer, U. (2003) Predicting reptile distributions at mesoscale: relation to climate and topography. *Journal of Biogeography*, **30**, 1233-1243.

Guisan, A. & Theurillat, J.-P. (2000) Equilibrium modeling of alpine plant distribution and climate change: how far can we go? *Phytocoenologia*, **30**, 353-384.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.

Guisan, A., Edwards, T.C. & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distribution: setting the scene. *Ecological Modelling*, **157**, 89-100.

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 modeling of plant species distribution. *Plant Ecology*, **143**, 107-122.

Hand, D.J. & Henley, W.E. (1997) Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society A*, **160**, 523-541.

Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002a) Ecological-niche factor analysis: how to compute habitatsuitability maps without absence data? *Ecology*, **83**, 2027-2036.

Hirzel, A., Hausser, J. & Perrin, N. (2002b) Biomapper 2.1 Manual. Laboratory for Conservation Biology, University of Lausanne, Lausanne, Switzerland [http://www.unil.ch/biomapper].

Hirzel, A.H., Helfer, V. & Métral, F. (2001) Assessing habitatsuitability models with a virtual species. *Ecological Modelling*, **145**, 111-121.

Houlder, D.J., Hutchinson, M.F., Nix, H.A. & McMahon, J.P. (2000) ANUCLIM User Guide, Version 5.1. Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia [http://cres.anu.edu.au/outputs/anuclim.html].

Jaberg, C. & Guisan, A. (2001) Modelling the distribution of bats in relation to landscape structure in a temperate mountain environment. *Journal of Applied Ecology*, **38**, 1169-1181.

Landolt, E. (1977) Ökologische Zeigerwerte zur Schweizer Flora. Veröffentlichtes Geobotanisches Institute ETH Stiftung Rübel, Zürich, Switzerland.

Lehmann, A., Overton, J.McC. & Leathwick, J. (2002) GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*, **157**, 189-208.

McCullagh, P. & Nelder, J.A. (1989) Generalized Linear Models, 2nd edn. Chapman & Hall, London, UK.

Manel, S., Dias, J.-M. & Ormerod, S.J. (1999) Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, **120**, 337-347.

Miller, R.I. (1986) Predicting rare plant distribution patterns in the southern Appalachians of the south-eastern USA. *Journal of Biogeography*, **13**, 293-311.

Moser, D.M., Gygax, A., Bäumler, B., Wyler, N. & Palese, R. (2002) Liste rouge des espèces menacées de Suisse. Fougères et plantes à fleurs. Office Fédéral de l'Environnement, des Forêts et du Paysage (OFEFP), Bern, Switzerland.

Myatt, M.M. (1987) Predicting the habitat geography of sensitive plants and community types. Conservation and Management of Rare and Endangered Plants (ed. T.S. Elias), pp. 173 -179. The California Native Plants Society, Sacramento, CA.

Ozenda, P. (1982) Les Végétaux Dans la Biosphère. Nathan, Paris, France. Peterson, A.T. (2001) Predicting species' geographic distributions based on ecological niche modelling. *Condor*, **103**, 599-603.

Rabinowitz, D., Cairns, S. & Dillon, T. (1986) Seven forms of rarity and their frequency in the flora of the British Isles. Conservation Biology (ed. M. Soulé), pp. 182-204. Sinauer Associates, Sunderland, MA.

Robertson, M.P., Caithness, N. & Villet, M.H. (2001) A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Diversity and Distribution*, **7**, 15-27.

Scott, J.M., Heglund, P.J., Haufler, J.B., Morrison, M., Raphael, M.G., Wall, W.B. & Samson, F. (2002) Predicting Species Occurrences. Issues of Accuracy and Scale. Island Press, Covelo, CA.

Sokal, R.R. & Rohlf, F.J. (1981) Biometry. Freeman, New York, NY.

Zaniewski, A.E., Lehmann, A. & Overton, J.McC. (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.

1.2

Testing the model-based sampling approach through simulation.

Robin Engler

This work is an extract of:

Using niche-based models to improve the sampling of rare species

Antoine Guisan[1], Olivier Broennimann[1], Robin Engler[1], Mathias Vust[1], Nigel G. Yoccoz[2], Anthony Lehmann[4] and Nicklaus E. Zimmermann[3].

This manuscript has been published in 2006 *Conservation Biology*, **20**, 501-511.

DOI: 10.1111/j.1523-1739.2006.00354.x

[1] Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland.

[2] Institute of Biology, University of Tromsø, 9037 Tromsø, Norway.

[3] Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland.

[4] Swiss Center for Faunal Cartography, Terreaux 14, CH-2000 Neuchâtel, Switzerland.

CONTRIBUTION TO THE PROJECT:
This is the subsection of the Guisan et al. (2006) paper that I worked on. I programmed, ran and analyzed the results of the simulations. I wrote the sections related to these simulations in Guisan et al. (2006). I also participated in the reviewing of the manuscript.

Preliminary note

The work shown below is an extract of "Using niche-based models to improve the sampling of rare species" by Guisan et al. (2006) with some additional material that was not included in the published manuscript. The complete version of Guisan et al. 2006 can be found in Annex 1 of this thesis.

Abstract (Guisan et al. 2006)

Because data on rare species usually are sparse, it is important to have efficient ways to sample additional data. Traditional sampling approaches are of limited value for rare species because a very large proportion of randomly chosen sampling sites are unlikely to shelter the species. For these species, spatial predictions from niche-based distribution models can be used to stratify the sampling and increase sampling efficiency. New data sampled are then used to improve the initial model. Applying this approach repeatedly is an adaptive process that may allow increasing the number of new occurrences found. We illustrate the approach with a case study of a rare and endangered plant species in Switzerland and a simulation experiment. Our field survey confirmed that the method helps in the discovery of new populations of the target species in remote areas where the predicted habitat suitability is high. In our simulations the model-based approach provided a significant improvement (by a factor of 1.8 to 4 times, depending on the measure) over simple random sampling. In terms of cost this approach may save up to 70% of the time spent in the field.

Keywords: efficiency, endangered species, *Eryngium alpinum*, habitat suitability maps, population discovery, predicted species distribution, prospective sampling.

Introduction

Quoting the words of McDonald (2004) in a book on the sampling of rare and elusive species (Poon and Margules 2004), "a rare population is one where it is difficult to find individuals because of small number, secretive and/or nocturnal behavior, or clamped distribution over large ranges, that is, a lot of zero occurs in the data". While McDonald had clearly animal species in mind, this definition can also been applied to plant species, possibly with an analogy between secretive/nocturnal behavior for animals and cryptic growth form/limited flowering time for plant species.

Of particular interest in this definition is its emphasis on the low spatial prevalence of rare species (i.e. "a lot of zero occurs in the data"). A direct result of low spatial prevalence is that, while standard field sampling methods such as simple or stratified random sampling, based on a simple combination of the main environmental gradients are generally working well with common species, they can be highly inefficient for rare and endangered species (Yoccoz et al. 2001, Rushton et al. 2004). For instance, in a sample of 550 plots surveyed in a random-stratified way based on the elevation, slope, and aspect (Randin et al. 2006), not one occurrence of the rare and endangered plant species *Eryngium alpinum* L. was recorded. Thus, using standard methods are likely to be neither time nor cost effective when attempting to sample rare species.

To increase sampling efficiency when working with rare species, one possibility that has been explored is to employ species distribution models (Franklin 1995, Guisan and Zimmermann 2000) to direct the sampling towards areas of potentially greater suitability (e.g. Poon and Margules 2004 and references cited therein, Edwards et al. 2005).

The basic idea behind model-based sampling is thus to use a habitat suitability model as stratification factor of a random-stratified sampling strategy (e.g. Wessels et al. 1998). In its simplest form, this stratification involves only two classes: suitable habitat and unsuitable habitat (or potential presence of the species and potential absence). Locations to be sampled are then randomly chosen in both the habitat that is projected as suitable and unsuitable. Given that rare species have by definition a low prevalence across the landscape, the number of sample locations corresponding to environmentally-favorable conditions that will be chosen with such stratification will be larger than if the sample locations had been chosen completely at random. In turn, visiting more sample locations associated with high environmental suitability is expected to result in higher likelihood of finding new populations of the target rare species.

Additionally, when implemented in this way, model-based sampling also meets the requirements of a true probabilistic sampling (i.e. all units have a non-null probability of being sampled and this probability is known), which is fundamental for ensuring unbiased population estimates (Yoccoz et al. 2001).

When monitoring a given species over several years, then model-based sampling can also be used as an iterative process (Guisan et al. 2006, Annex 1). A first model is fitted with initial available data. These data typically derive from information on monitored populations or records of species that are extracted from a computer data bank, such as those developed in botanical conservatories, museums, or biological data centers (Graham et al. 2004). Spatial predictions derived from this initial model are then used to stratify a first field sampling (i.e.,

model-based sampling). Data gathered during this first sampling campaign serve to update the training data set and fit improved models that are used in turn to direct the next sampling campaign, and so on iteratively over several field seasons (see Figure 1 of Guisan et al. 2006 in Annex 1). The higher prediction strata are expected to reveal those occurrences of suitable environmental conditions for the species where undetected or newly established populations, not recorded in the initial data set, might prevail. This iterative process can be repeated during the same field season or over consecutive seasons, and its success can be measured by the number of new occurrences found at each stage.

As field-work is time and resource consuming, testing the long-term potential of this iterative model-based sampling approach is difficult to do in practice, and is thus best done through simulation. A further advantage of using a simulation approach with a virtual species (e.g. as in Hirzel and Guisan 2002) is that the "true" distribution of the test species is fully known, an information that is difficult to obtain for any real species.

Here I theoretically test the iterative model-based sampling approach by running simulations for a virtual model-based sampling procedure for a virtual species over a period of 10 years. At each iteration, I measure the improvements in terms of model accuracy (i.e., match between modeled and true virtual distribution) and measure the number of new populations found during virtual field work. I run different simulations to test the sensitivity of the model-based sampling approach to initial conditions and sampling effort. Finally, I compare the results to those obtained through a simple complete random sampling strategy.

Methods

The true distribution of the virtual species in the study area (whole of Switzerland) was derived from a generalized additive model (GAM; Hastie and Tibshirani 1986) calibrated for *Eryngium alpinum* using 92 presences and 2380 absences. The details of the environmental variables and occurrence data used to calibrate this model are given in Guisan et al. (2006) in Annex 1 of this Thesis. This ensured that our simulations would remain as close as possible to a real situation (same type of distribution, same study area).

To turn the probabilistic map into a binary presence-absence map, we used a cut-off value of 0.8: pixels where the GAM projected values above that threshold where classified as being part of the virtual species' distribution. This procedure yielded a spatial distribution restricted to about 5% of the study area and was considered as being the true distribution of our virtual species.

Knowing this "true distribution" is essential to carry-out the virtual field sampling, which is then simply done by checking whether the sampling location corresponds to a presence or an absence of the virtual species on the true distribution map.

For computational efficiency reasons the simulations were not run over the entire Swiss landscape but on a subset of 500'000 pixels selected in a random-stratified way across the main environmental gradients. This subset of pixels accurately reflected the prevalence of the different habitat suitability classes of our virtual species.

The iterative, model-based, random-stratified sampling procedure was simulated as follows (see also flow chart in Figure 1):

1. An initial data set of *N* presences and *N* absences was randomly selected among the pools of 92 presences and 2380 absences used to calibrate the "true distribution" of our virtual species. Hereafter we refer to this data set as the training data set.

2. Using the training data set, a GAM was fitted and projections of habitat suitability were made across the 500'000 randomly chosen pixels (i.e. the subset of the entire Swiss landscape).

3. The probabilistic projections were reclassified into binary projections (i.e. habitat is suitable or not) – as required by the model-based sampling procedure. This conversion was done using so that the projected suitable habitat was as small as possible while still encompassing all of the observed occurrences of the species.

4. The agreement between the binary predictions map and the true distribution of the virtual species was evaluated using the kappa coefficient (Cohen 1960). This was done to provide a measure of model improvement over time (i.e. over the successive iterations), reflecting a convergence, or non-convergence, toward the virtual species' true distribution.

5. An equal number *N* of virtual sample sites were selected randomly within the habitat modeled as suitable and the habitat modeled as unsuitable (i.e. the two sampling strata). To avoid spatial autocorrelation we did not allow the selection of new sample sites within a radius of 500 m of any already sampled site.

6. The fieldwork was simulated by checking the selected virtual sampling locations against the true distribution of the virtual species, allowing to assign a (virtual) true presence or absence to each sample site. At this stage we recorded the number of new occurrence sites found at each iteration. All new presence and absence data that were collected during the virtual sampling were then added to the training data set.

7. Steps 2 to 5 were repeated 10 times (i.e., simulating a 10-year survey). Because the selection of the new sample sites involved randomness, we repeated the entire procedure 20 times to quantify the related uncertainty (error bars in Figure 2).

I ran this simulation procedure for three different experiments: sample30, sample60, and sample120. Numbers indicate the initial number of observations and the number of sites sampled at each iteration. For instance "sample 30" means that the data at the start of the simulation (Step 1) was composed of 15 presence and 15 absence data and that 15 virtual sampling sites within the "predicted presence" strata and 15 sites within the "predicted absence" strata were visited.

Additionally, control simulations in which the new sample sites were chosen completely at random in the landscape, rather than on a model-based stratification basis, were run for each scenario (dashed lines, Figure 2). I also repeated these control simulations 20 times. Differences between model-based and full-random-sampling simulations were assessed with paired t-tests or Wilcoxon signed rank tests when normality was not satisfied. When p-values were > 0.05, the difference between the two curves was not significant.

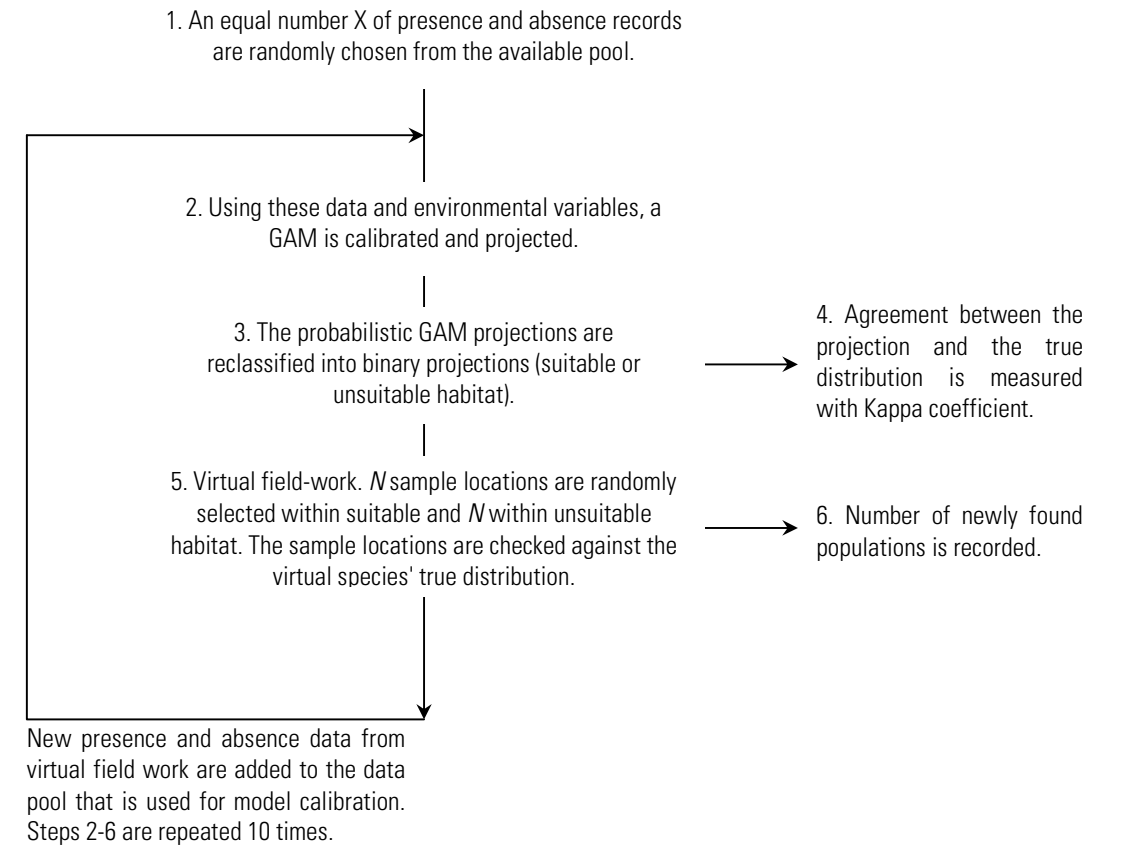


Figure 1. Flow chart of simulations run to test the model-based sampling approach. The entire procedure was repeated 20 times for each simulation.

Finally, to test if a bias in the initial set of occurrence and absence data had a significant influence on the outcome of model-based sampling, a "Sample60 biased" simulation was also run. In this simulation, the initial presence data were not chosen randomly but in an environmentally biased fashion, i.e. only locations with similar environmental characteristics were selected.

Results

In all three scenarios the number of new presences found over 10 iterations was always much higher when using the model-based stratification approach (about four times in all simulations; Figure 2, upper graphs).

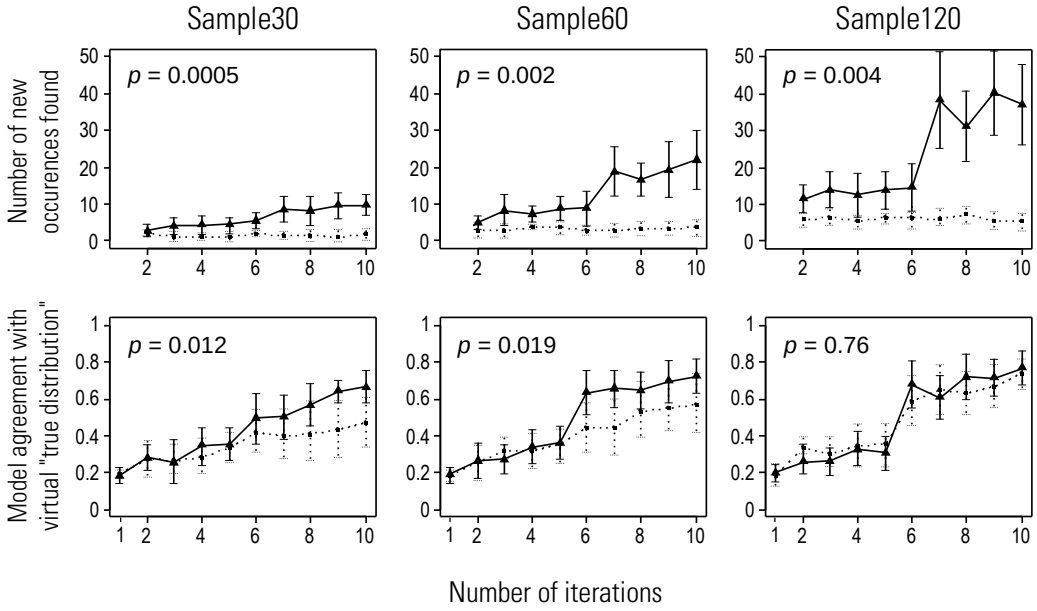


Figure 2. (Figure 3 of Guisan et al. 2006) Average (±SE) number of new occurrences of Eryngium alpinum at each iteration (upper graphs), and the kappa agreement between the predictive model built at each iteration and the true distribution of the virtual species (lower graphs). Solid lines represent simulation with model-based stratification when selecting the new locations to sample. Dashed lines are the control simulation, where the new points to be sampled were chosen completely at random. The horizontal series of graphs correspond to increasing number of sample points visited at each iteration (30, 60, 120). The p values are only indicative because they depend on the number of simulations. Each point on a graph was obtained as an average of 20 runs. The field work of iteration i + 1 is done using the model of iteration i. This explains why the "number of new occurrences found" starts at iteration 2.

For instance, at a sample size of 30 ("Sample30"), only 6 iterations were needed with the model-based sampling to double the initial number of presences (i.e. to pass from 30 to 60), whereas nearly 20 iterations would have been needed with simple random sampling. In this case, the model-based approach was 3.3 times more efficient, corresponding to a net gain in time of 70%. At a sample size of 60 ("Sample60"), the model-based approach was still 2.3 times more efficient (considering the time needed to double the number of occurrences), allowing a 55% gain of time compared with random sampling. In all experiments, the cumulative number of new occurrences found by model-based sampling after 10 iterations was always about 4 times greater than those found by a simple random sampling. Averages and standard errors of the adequacy between the predictions and the true distribution of the virtual species are also reported for all simulations (Figure 2, lower graphs). Except for the sample120 scenario, where no significant difference was found, the model-based sampling

approach led, on average, to a predicted distribution that had higher agreement with the true distribution of the virtual species than that obtained through complete random sampling.

Even when the initial available data was environmentally biased ("Sample60 biased" simulation), the model-based sampling approach performed almost as good as its unbiased counter part (Figure 3). The number of new populations found is almost identical between the biased and unbiased simulation and the convergence of the model towards the true distribution of the virtual species is only a little slower during the first two iterations. After this the initial bias appears to be corrected.

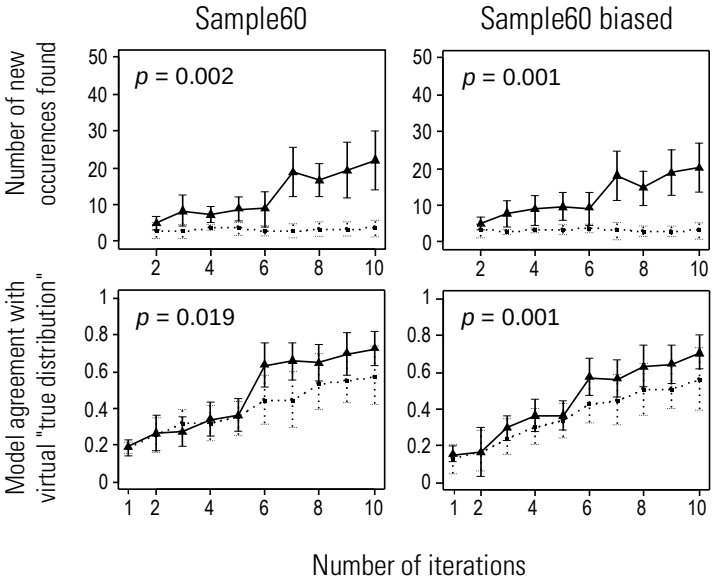


Figure 3. Average (±SE) number of new occurrences of Eryngium alpinum at each iteration (upper graphs), and the kappa agreement between the predictive model built at each iteration and the true distribution of the virtual species (lower graphs). Solid lines represent simulation with model-based stratification when selecting the new locations to sample. Dashed lines are the control simulation, where the new points to be sampled were chosen completely at random.

Discussion

The simulations provided theoretical evidence that using model-based sampling rather than a simple random sampling (control) yields both a greater number of presences and a higher adequacy in terms of model predictions (figure 2 and 3). Although such improvement over a random sampling is not surprising, in fact probably no one would sample rare species randomly in the field, the simulation approach nevertheless provides a standardized and objective measure of improvement. The simulation results also show how the iterative

approach proposed in Guisan et al. (2006, Annex 1), has the potential of increasing model quality and the knowledge of a species' distribution (i.e., the model predictions converge towards the species' true distribution). Taken together, these results thus corroborate, from a simulation perspective, the findings from others that have shown the efficiency of model-based sampling in practice by doing actual field-work (Engler et al. 2004, Edwards et al. 2005, Guisan et al. 2006, Le Lay et al. *submitted* – section 1.3 of this thesis).

Obviously, the difference between model-based sampling and completely random sampling will depend on the spatial prevalence of the target species in the landscape, and if a species is very common (~50% spatial prevalence) then model-based sampling will become an unnecessary complication.

The improvements yielded by the model-based approach over complete random sampling are also mostly true below a certain sample size. With 120 sites (or more) sampled at each iteration, the model-based approach did not converge significantly faster toward the true distribution of the virtual species than the simple random sampling. However, the number of new presences found remained much greater (nearly four times). With a sample size of 30 and 60 sites visited at each iteration, the improvements are very apparent, with a significantly greater number of presences found at each iteration. Thus, the lower the number of sites sampled at each iteration, the more valuable the use of the model-based approach becomes. This has direct implications for reducing surveys costs in nature conservation applications.

Acknowledgments

To B. Baümler at the Swiss Floristical Center in Geneva for making the species data available, as well as to C. Trippi, and Y. Naciri-Graven for additional species data.

References

Cohen J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37-46.

Edwards, T. C. J., R. Cutler, N. E. Zimmermann, L. Geiser, and J. Alegria. 2005. Model-based stratifications for enhancing the detection of rare ecological events: lichens as a case study. *Ecology* **86**, 1081-1090.

Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* **41**, 263-274.

Franklin J., 1995. Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19** (4), 474-499.

Graham C.H., Ferrier S., Huettman F., Moritz C., Peterson A.T., 2004. New developments in museum-based informatics and applications in biodiversity analysis, *Trends in Ecology and Evolution*, **19** (9), 497-503.

Guisan A. and Zimmermann N.E., (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.

Guisan A., Broennimann O., Engler R., Vust M., Yoccoz N. G., Lehmann A. and Zimmermann N. E., 2006. Using niche-based models to improve the sampling of rare species, *Conservation Biology*, **20** (2), 501-511.

Hastie T. and Tibshirani R., 1986. Generalized additive models. *Statistical Sciences*, **1**, 297–318.

Hirzel A. and Guisan A., 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331-341.

McDonald L.L., 2004. Sampling Rare Populations. In Poon E.L. and Margules C.R. (Eds.), 2004. Searching for new populations of rare plant species in remote locations. Pages 189–207 in W. L. Thompson, editor. Sampling rare or elusive species. Island Press, Washington, D.C.

Poon E.L. and Margules C.R., 2004. Searching for new populations of rare plant species in remote locations. Pages 189–207 in W. L. Thompson, editor. Sampling rare or elusive species. Island Press, Washington, D.C.

Randin C.F., Dirnbock T., Dullinger S., Zimmermann N.E., Zappa M. and Guisan A., 2006. Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**,1689-1703.

Rushton S.P., Ormerod S.J., and Kerby G., 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology*, **41**, 193-200.

Wessels K.J., van Jaarsveld A.S., Grimbeck J.D. and van der Linde M.J., 1998. An evaluation of the gradsect biological survey method. *Biodiversity and Conservation*, **7**, 1093-1121.

Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, **16**, 446–453.

1.3

Using habitat - suitability models enhances chances to find rare species in the field.

Gwenaëlle Le Lay[1], Robin Engler[1], Erika Franc[1], and Antoine Guisan[1]

This manuscript will be submited to *Ecography*.

[1] Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland.

CONTRIBUTION TO THE PROJECT:
I developed the original idea in collaboration with GLL and AG. I prepared the data and performed the modeling in collaboration with EF and participated to the field-work. I also participated in the writing and reviewing process.

Abstract

Identifying the geographic distribution of populations is a basic, yet crucial step in many fundamental and applied ecological projects, as it provides key information on which many subsequent analyses depend. However, this task is often costly and time consuming, especially for rare species for which most sampling designs generally prove inefficient. At the same time, rare species are those for which distribution data are most needed for their conservation to be effective. To solve this paradigm, model-based sampling (MBS) can use predictions from habitat suitability models to increase field sampling efficiency, potentially supporting the discovery of new populations. Here, we thoroughly test the efficiency of MBS by conducting a large field campaign, assessing whether MBS can improve the detection rate of three rare and five common plant species. For each species, we gathered existing occurrence data and used them in combination with environmental variables to produce models and derive habitat suitability maps following an ensemble forecasting framework designed to better outline the highly suitable areas. We tested the efficiency of MBS and the accuracy of our models by sampling 240 sites in the field (30 sites × 8 species). Across all species, the MBS approach proved largely efficient. It particularly led to the discovery of six new populations of one rare plant, increasing chances to find this species from 0 to 50%. For common species, MBS doubled the new population discovery rates as compared to random sampling. We also demonstrate how combined models can predict the species distribution more accurately than individual modeling methods. We conclude that conservation of rare species could be more efficient if MBS was implemented in practice. More generally, we recommend habitat suitability models to be used in support of conservation plans.

Keywords: Habitat suitability, spatial predictions, species distribution, field validation, sampling, rare plants, ENFA, GAM.

Introduction

Identifying where species occur in the landscape is of primary importance for many conservation issues, such as the management of pest species (Lodge et al. 2006), or the selection of protected areas (Margules & Pressey 2000). Unfortunately, information on geographic distribution is often incomplete, especially for rare species (Pulliam & Babbitt 1997). Due to their rarity, these species are usually absent from most of the sampled areas and sampling designs have thus to be adapted to them (Yoccoz et al. 2001). Finding rare species in the field is often a very difficult task, requiring a large amount of human, time and funding resources, which are generally unavailable.

Habitat suitability models are powerful tools (e.g. Guisan & Zimmermann 2000; Hirzel & Le Lay 2008) which can efficiently support conservation practices (Côté & Reynolds 2002). These models relate data about species presence/absence or abundance, coming from a training dataset, to environmental variables. The calibrated models can then be projected onto the whole study area, producing a map that predicts how suitable each mapping unit is for the target species. One important recent application of habitat suitability models in biological conservation is forecasting where rare and endangered species (Griffin et al. 2003; Edwards et al. 2005; Aitken et al. 2007) or rare species richness (Miller 1986) are most likely to be observed in the landscape. More technically, these models have been proposed to improve the design and efficiency of field sampling campaigns, by using classes of predicted habitat suitability as a stratifying factor (e.g. Guisan et al. 2006a). In comparison with traditional approaches, such as random or regular sampling (Hirzel & Guisan 2002), model-based sampling (MBS) is expected to support an increased rate of detection of new populations in the landscape, therefore contributing to reduce sampling cost. Some studies partially tested the efficiency of MBS (Boetsch et al. 2003; Engler et al. 2004; Edwards et al. 2005; Guisan et al. 2006a; Aitken et al. 2007), suggesting the high potential of such methods. Nevertheless, the MBS approach still requires more testing in the field and methodological improvements before being implemented routinely in conservation practices, in particular assessments involving multiple species of various ecological types, detectability level and abundances. Modeling methodology could be improved to better highlight areas where a species has higher occurrence probability. One particular area where improvements could be made is the utilization of the available species occurrence data, such as public or museum databases, which knowledge is often insufficiently valorized. Occurrence data often come from various sources, with variable spatial and temporal accuracies (Graham et al. 2004). Using a coarse resolution permits to use all observations, but at the risk of losing spatial accuracy in the description of the species-environment relationship. At a fine resolution, some coarse observations have to be discarded, possibly resulting in a reduced sample size and a loss of important ecological information (Engler et al. 2004). One solution to this dilemma is to adopt a combined framework, in which models at different scales are combined to yield a single final prediction. To our knowledge, such approach has never been applied to model the distribution and improve the sampling of rare species.

Finally, many papers highlighted difficulties in choosing a particular habitat modeling technique. Even if some techniques seems to have better predictive power than others (Moisen & Frescino 2002; Elith et al. 2006), the choice of the best technique can nevertheless remain difficult, owing to the ranking of techniques possibly changing depending on the species. A better alternative seems to combine several modeling techniques into an ensemble of predictions (Araújo & New 2007), as it should minimize the risk of prediction failure that can arise from individual models. Here we use such an ensemble forecasting

approach to combine models built at different resolutions and with different modeling techniques.

In this study, we conducted an extensive fieldwork to thoroughly test such combined model-based sampling (MBS) approach on several plant species, both rare and more common, with the latter serving as a methodological control. By sampling these species in each of the predicted suitability classes, we test whether our combined approach improves model quality as well as the discovery rate of new populations in the field. The test is based on a large field survey conducted posterior to the model fitting, throughout a study area in the Swiss Alps during an extended summer season (May-October 2005). Our work thus provides the first large independent test of the MBS approach.

Methods

Study area

The study area is located in the western Swiss Alps (centered on 7°10'E, 46°20'N; Fig. 1), overlapping two of the Swiss biogeographic regions: "Northern Alps" (NA) and "Western Internal Alps" (WIA) (Gonseth et al. 2001). Elevation ranges from 372 m to 3237 m a.s.l. and the area totals 2054 km².

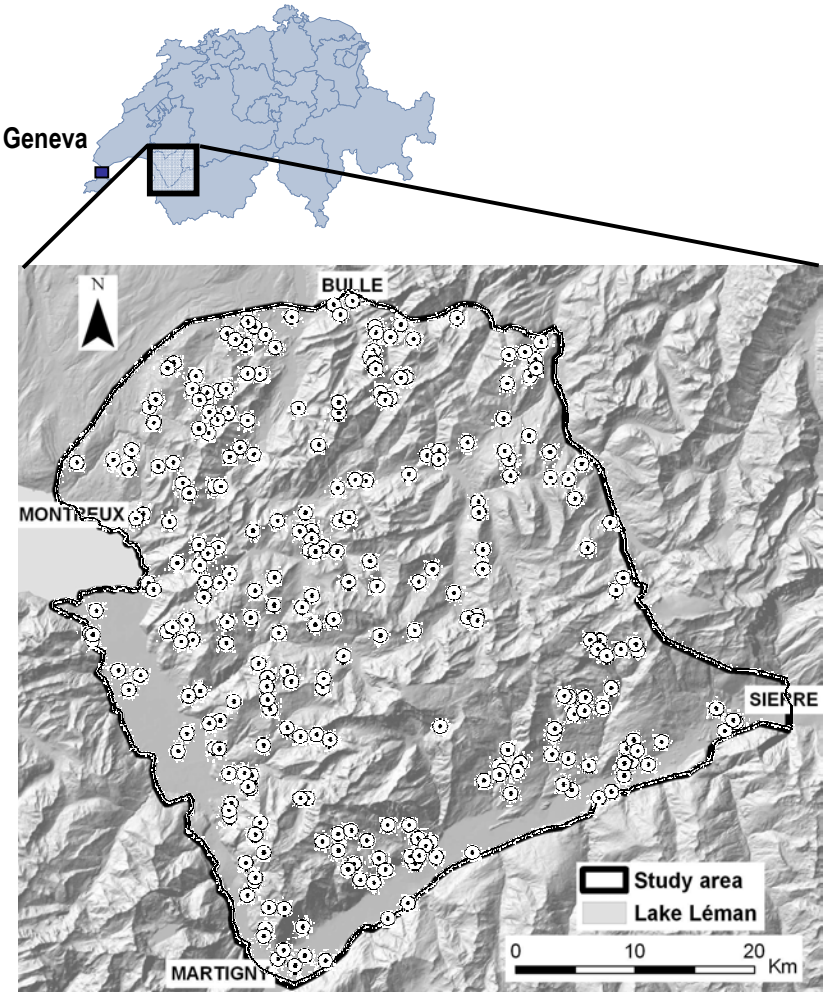


Figure 1. Localization of the study area in the Western Swiss Alps. White points indicate locations sampled during fieldwork to test the model-based sampling (MBS) method.

The climate in the WIA is quite warm and dry, whereas the NA are colder and have more precipitations. The WIA are notably characterized by vineyards and orchards, while pastures are mainly found in the NA, up to the highest altitudes. Forests are quite abundant overall but absent at the lowest elevation. Human activities vary in type and intensity throughout the study area. Below 1000 m, towns, roads and agricultural areas have highly reduced and fragmented natural ecosystems. Along the altitudinal gradient, the sequence of vegetation belts is typical for the Western Alps (see Randin et al. 2006). Vegetation is influenced by human land use: pasturing is common in this region and skiing resorts are developed in several sites. Some parts of the study area are protected.

Study species

Rare species were selected using four criteria. The species had to (i) be mentioned as threatened in the IUCN red lists of Switzerland, (ii) have enough available occurrence data to allow model-fitting (> 10 occurrences), and (iii) be easily identified in the field. Three species were selected (Table 1): *Cypripedium calceolus L.*, *Eryngium alpinum L.* and *Scorzonera laciniata L.*

We also selected five common species (Table 1) based on their ecological similarities with the selected rare plants (i.e. habitat types, phenology) to serve as a methodological control. They present various abundances in the study area (Table 1), but have all a sufficiently high prevalence for successful testing of the MBS approach.

Table 1. Study species, their main habitat characteristics and general reproductive traits. Number of presence and absence points in the study area at both spatial accuracies, i.e. < 50 m and up to 750 m, are given in the leftmost columns.

	Name	Habitat	Morphology	# Presences		#Absences	
				50m	1km	50m	1km
Rare species	*Cypripedium calceolus*	deciduous/mixed woodland, open scrub, alpine meadows 300-2000m a.s.l.	15-50 cm high, large rounded leaves. Groups of several individuals. Yellow flowers of 3-4 cm high, in May-June	14	25	553	388
	Eryngium alpinum	avalanche corridors, hayfields 1400-2100m a.s.l.	30-70 cm high, large leaves. Individuals dispersed in the vegetation. Blue flowers (capitule 6 cm high) in August	43	79	546	362
	Scorzonera laciniata	vineyards, road banks 400-1400m a.s.l.	10-40 cm high, fine linear grey-green leaves. Plants often grouped in small populations. Yellow flowers, ~1.5 cm high, in June-July.	31	81	550	377
Common species	*Anthyllis vulneraria*	dry meadows, open forest <1600m a.s.l.	15-40 cm high. White-yellow flowers of 3-4 cm wide, from May to September	199	200	2272	531
	Astrantia major	mountainous meadows, clearings <2400m a.s.l.	30-90 cm high. White-pink flowers (2-3 cm large) from June to September.	479	484	1879	338
	Briza media	meadows, oligotrophic pastures 700-2400m a.s.l.	20-50 cm high. A few spikelets of 5-7 mm high, from June to September.	206	212	2166	500
	Heracleum sphondylium	Meadows, megaherbs <2400m a.s.l.	50-150 cm high, large leaves. White flowers grouped in large (about 10 cm wide) umbels, from June to September.	892	909	1686	1673
	Pulsatilla alpina	Meadows 1300-2400m a.s.l.	20-50 cm high. White flowers, 3-6 cm wide, from May to July.	195	197	1817	475

Modeling framework

Our modeling framework consisted of four steps (Fig. 2). For each species we: (1) prepared species and environmental data, (2) modeled habitat suitability using two different techniques - generalized additive model (Hastie & Tibshirani 1986), requiring presence/absence data, and ecological niche factor analysis (ENFA, Hirzel et al. 2002), requiring only presence data - at two different spatial resolutions (50 m and 1 km), (3) reclassified the predictions into binary presence/absence values, and (4) combined the four reclassified maps (i.e. resulting from the two techniques and two resolutions) to produce a single final prediction map.

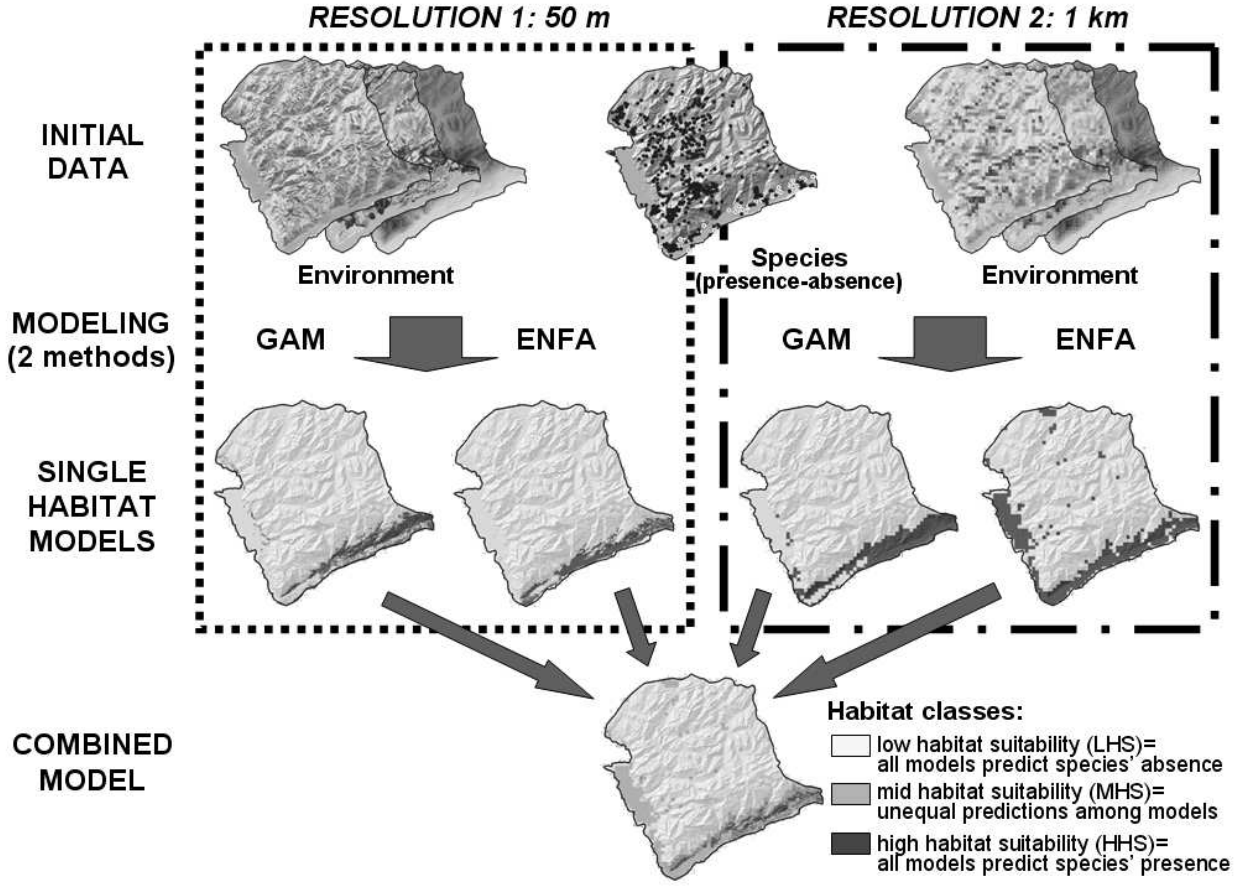


Figure 2. Modeling framework: 1) Environmental and species data are prepared at two spatial resolutions: 50 m and 1 km. 2) Habitat suitability (HS) maps are created for each species at both resolutions using two modeling methods: GAM (Generalized Additive Models) and ENFA (Ecological Niche Factor Analysis). 3) HS maps are reclassified into binary prediction maps, i.e. habitat is either favorable or not. 4) For each species, the four binary HS maps are combined to provide a final map comprising three categories: low habitat suitability (LHS), medium habitat suitability (MHS), and high habitat suitability (HHS).

Species data

The two main data sources used for this study were (i) presence-only data, obtained from public observations stored in the Swiss floristic database (CRSF; http://www.crsf.ch) and from the literature (Droz 1994), and (ii) presence-absence data, obtained from exhaustive vegetation plots (Clot & Delarze 1998; Randin et al. 2006). The rare species data mainly came from the first source, which is rather heterogeneous, with large variations in spatial (1 m to 1 km) and temporal (late 19th century to 2005) resolution. We only considered observations from 1950 to 2005, and used only those absences in the database that fell within the growing and flowering periods of the target species. Data for the common species came mainly from the exhaustive vegetation plots, with a spatial accuracy finer than 25 m.

To deal with the data's variability in spatial accuracy, we divided observations into two classes: accuracy ≤ 50 m and accuracy ≤ 750 m, the latter class including records from the former. The first dataset was used to develop models at a resolution (pixel size) of 50 m, whereas the second supported models at a 1 km resolution.

Environmental variables

We chose environmental variables that reflected the main environmental characteristics of the study area and the species' ecological and ecophysiological requirements (Table 2). Topo-climatic variables were often demonstrated to be important parameters for modeling plant distribution in mountainous regions (Zimmermann & Kienast 1999; McKenzie et al. 2003). We also used a normalized difference vegetation index (NDVI) as surrogates for vegetation cover, and different geological classes as potential indicators of soil characteristics (Table 2).

Variables were prepared in ArcGIS (ESRI, Redlands, USA) at 50 m and/or 1 km resolutions, except for topographic position and NDVI that were only prepared at 50 m resolution (uninformative at 1 km), and edaphic factors that were only available at 1 km resolution (Table 2). Correlation among variables never exceeded 0.5 at 50 m resolution and 0.7 at 1 km resolution.

Table 2. Environmental variables used for species distribution modeling. The two last columns indicate the resolution these data were used.

Variable name	Description	Range in the study area	Source	Computed resolution	
				50m	1km
SLOP	Slope (in degrees)	[0 : 81.5]	Calculated from Digital Elevation Model (DEM) at 25 m resolution (Zimmermann & Kienast 1999)	x	x
TOPO	Topographic position, i.e. relative lower or higher elevation position of a central cell of a moving window in comparison to its surrounding cells.	[-2337.5 : 1953]		x	
SRAD	Mean of short wave radiations (direct and diffuse) per day from May to September (in KJ/m²/day).	[3056.2 : 29558.2]		x	x
MIND	Mean monthly moisture index from May to September (in (1/10mm)/month). This index comes from subtraction of the potential average daily evapotranspiration from the mean daily precipitations	[-1195 : 1820.5]	(Zimmermann & Kienast 1999)	x	x
NDVI	Index combining the red and the near-infrared bands of the image. It reflects the photosynthesis activity of the vegetation.	[-0.557 : 0.759]	Calculated from Landsat ETM image 10/2001	x	
CALC	Soil richness in limestone	[0;1;2;3;4;5]		x	x
PERM	Soil permeability	[0;1;2;3;4;5]		x	x
GTC1	Quaternal mobile deposits	[0;1]	Expert assessment from geological data (Lehmann & Ayer 2004)		x
GTC2	Detritical rocks fines and medium	[0;1]			x
GTC3	Limestone	[0;1]			x
GTC4	Fine detritical rocks (clays)	[0;1]			x
GTC5	Limestone magnesic and evaporites	[0;1]			x

Habitat suitability models

Generalized Additive Models (GAMs) were fitted, for each species, with the available presence and absence data, using the GRASP package (Lehmann et al. 2002). Models were selected with a bidirectional stepwise procedure based on the Akaike Information Criterion (AIC; Akaike 1973). As most species had many more absence than presence records, absences were weighted to achieve equal prevalence, ensuring that model calibration would not be overly driven by the more numerous absence records. Indeed, our goal was to focus on the potential highly suitable areas and, in case of species that may not be at equilibrium with their environment, taking into account fallacious absences can hamper models to correctly predict these highly suitable areas (Cianfrani et al. submitted). Models were evaluated using cross-validated AUC (cvAUC), for which values from 0.7 to 0.9 indicate useful predictions, and 0.9 to 1 indicate good predictions (Swets 1988). The fit of models was also assessed from their adjusted deviance (*adj. D^2*), an index proportional to a model's explained deviance and corrected for the effective number of degrees of freedom used (Guisan & Zimmermann 2000). *Adj. D^2* ranges from 0 (no fit) to 1 (perfect fit).

Ecological niche factor analysis (ENFA) models were fitted using the Biomapper software (Hirzel et al. 2006a). As this software requires inputs to be raster maps, we constructed quantitative maps that took into account all the species presences falling into the same raster cell. We used two criteria: (1) when both presence and absence occurred in a same cell, presence information prevailed, and (2) for the rare species, the number of presences aggregated in a cell was kept as a quantitative measure. This methodological choice was again guided by the goal of our study, i.e. obtaining models especially focusing on the highly suitable areas, thus keeping as much as possible the information on the species presence. In the case of common species, as presence points are numerous enough for modeling, we aggregated the points falling in the same 1 km pixel. We evaluated ENFA predictions using a cross-validated continuous Boyce index (see Hirzel et al. 2006b). This index ranges from -1, for models inversely predicting the species occurrences, to 1 for excellent models.

Reclassifying model predictions

Each of the four models (2 techniques × 2 resolutions) that we fitted for each species yielded projections along a gradient of habitat suitability. Using a reclassification threshold, we transformed these continuous projections into binary maps: 0 indicating unsuitable habitat, and 1 indicating suitable habitat. For the GAM-based projections, this threshold was chosen to maximize the weighted Kappa (Cohen 1968). For the ENFA-based models, we used the characteristics of the predicted *vs.* expected ratio curves derived from the cross-validation procedure (Boyce index), i.e. keeping as many occurrences of a species as possible in the suitable area while keeping the surface of this area as small as possible (see Hirzel et al. 2006b).

Combining predictions

For each species, we combined the four binary projections following an ensemble forecasting approach (Araújo and New, 2007). The combined maps were reclassified into three habitat suitability classes: low habitat suitability (LHS) where all individual models predicted 0; medium habitat suitability (MHS) where predictions disagree, and high habitat suitability (HHS), where all four individual models predicted 1. This combination aims at targeting the areas of highest habitat quality (i.e. the HHS), increasing chances to find the species, while shifting the marginal habitats and potential modeling uncertainties in an intermediate class (MHS). All suitability maps were also filtered to avoid sampling in evidently unsuitable areas (e.g. lakes and glaciers).

Model-based field sampling (MBS)

We used the combined models to stratify the field sampling. For each species, ten sampling locations were randomly chosen within each habitat class (LHS, MHS and HSS), yielding a total of 30 sampling locations per species. The sampling was further constrained to fall within the altitudinal range of the species (see Table 1). From May to October 2005, we visited each sampled location and recorded the presence or absence of our eight species in a 50 × 50 m quadrat. The position of quadrats were spatially adjusted to be entirely contained within a

single habitat class. We also avoided overly artificial or disturbed sites, such as roads or highly pastured areas.

Field evaluation of MBS

For each species, we compiled all the presences and absences recorded in the 240 plots (8 species × 30 plots) visited during fieldwork. From the 30 sample points of each species, we calculated the ratio of species presences found in the HHS class and the species absences found in the LHS class, i.e. the positive predictive power (PPP) and the negative predictive power (NPP), respectively, of the combined model (Fielding & Bell 1997).

For each species, we evaluated the efficiency of our MBS by comparing it to a random sampling procedure simulated a posteriori from our 240 field plots. For each species, we calculated observed prevalences from (i) 10 random points taken from the 240 sampled locations (whole study area), (ii) 10 random points from any sampled site located in the species' HHS class, and (iii) the 10 points targeted with our MBS in the species' HHS class.

For each 10-point sample, we also calculated a weighted Kappa (Cohen 1968). The Kappa statistic indicates how observed presence-absence data agree with the presence-absence data predicted by the model, and ranges from 0 (no agreement) to 1 (full agreement). Unlike the commonly used Kappa (Cohen 1960), the weighted Kappa index weights the contingency table so that the most prevalent category (i.e. presences or absences) does not bias the result. We calculated weights as to equal total sampling in each habitat type.

Finally, to allow comparison of models based on their global predictions, we also calculated the AUC index on the individual habitat suitability maps with continuous values. AUC has the advantage of being independent of the choice of a reclassification threshold (Fielding & Bell 1997). However, it was not possible to calculate AUC for the combined models, as these were based only on reclassifications (i.e. binary maps).

The entire evaluation approach was repeated 1000 times to obtain uncertainty estimates.

Results

Fit and evaluation of models

Models were successfully fitted with both techniques and at both resolutions for all eight species. Despite the small number of presences available for rare species, GAM explained between 51% and 82% of the deviance (Table 3). Cross validation (cvAUC) values ranged between 0.83 and 0.97 (Table 3), indicating useful to good predictions (Swets 1988). ENFA models obtained values for the continuous Boyce index (Hirzel et al. 2006b) between 0.48 and 0.75 for the common species at both resolutions, and between 0.23 and 0.24 for the models of rare species at the 50 m resolution and between 0.38 and 0.72 at the 1 km resolution. While topo-climatic variables proved overall to be useful predictors to explain the distribution of our eight species, their success varied across species, spatial resolution and modeling technique. Overall, moisture index (MIND), slope (SLOP) and solar radiations

(SRAD) appeared to be the most efficient variables for explaining the distribution of all species. No particular trend in environmental variable usefulness was found between resolutions (50 m *vs*. 1 km) or for common species *vs* rare species.

Spatial predictions

For the common species, the combined predictions yielded fairly small HHS and LHS areas, i.e. areas where all four model predictions agreed, suggesting disagreement between techniques and resolutions for these species (Fig. 3b). Conversely, two of the rare species, *Eryngium alpinum* and *Scorzonera laciniata*, had quite small predicted MHS areas (Fig. 3b), suggesting that the models gave more similar predictions for the distribution of the suitable habitat for these species. When considering LHS and HHS only, predictions for the rare species are dominated by LHS values, whereas predictions for the common species are dominated by HHS values, highlighting respectively the observed rarity and commonness of these species.

Table 3. Fit of the GAMs (*adj. D^2*), cross-validation indices of the ENFA and GAMs, and validation with independent data (independent validation) obtained from field work. For cross validation of the ENFA models, the Boyce values are for the continuous index (window size=20), values in brackets indicate standard deviation.

	Species	Map resolution	Cross validation			Independent validation				
			GAM		ENFA	GAM		ENFA		Combined model
			D2	cvAUC	Boyce	WKappa	AUC	WKappa	AUC	WKappa
Rare species	Cypripedium calceolus	50 m	0.51	0.83	0.24 [0.20]	-	-	-	-	-
		1 km	0.58	0.85	0.48 [0.39]	-	-	-	-	-
	Eryngium alpinum	50 m	0.75	0.97	0.23 [0.33]	-	-	-	-	-
		1 km	0.69	0.95	0.38 [0.46]	-	-	-	-	-
	Scorzonera laciniata	50 m	0.82	0.97	0.24 [0.39]	0.17	0.98	0.26	0.95	0.31
		1 km	0.73	0.96	0.72 [0.16]	0.15	0.98	0.09	0.89	
Common species	Anthyllis vulneraria	50 m	0.27	0.81	0.62 [0.18]	0.03	0.54	0.34	0.72	0.44
		1 km	0.27	0.80	0.48 [0.25]	0.27	0.67	0.27	0.71	
	Astrantia major	50 m	0.08	0.67	0.62 [0.18]	0.30	0.74	0.30	0.70	0.59
		1 km	0.17	0.71	0.70 [0.21]	0.37	0.79	0.27	0.71	
	Briza media	50 m	0.19	0.76	0.53 [0.27]	0.40	0.72	0.16	0.60	0.64
		1 km	0.11	0.68	0.48 [0.30]	0.40	0.76	0.16	0.65	
	Heracleum sphondilium	50 m	0.03	0.56	0.75 [0.22]	0.22	0.64	0.18	0.64	0.38
		1 km	0.07	0.64	0.73 [0.11]	0.24	0.67	0.20	0.67	
	Pulsatilla alpina	50 m	0.40	0.87	0.65 [0.27]	0.33	0.87	0.26	0.81	0.51
		1 km	0.46	0.89	0.75 [0.24]	0.35	0.85	0.24	0.74	

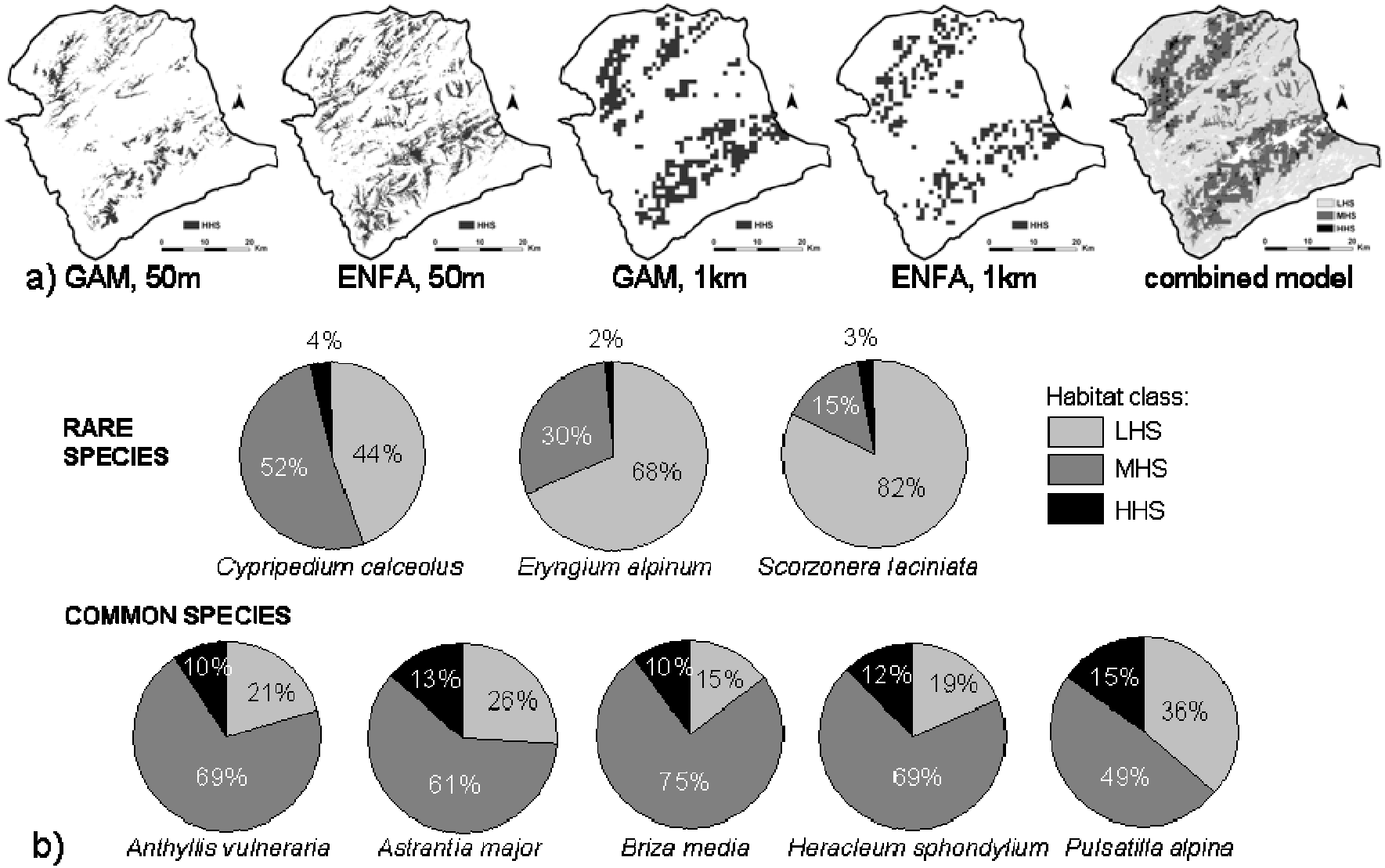


Figure 3. Discrepancies between model predictions illustrated by a) maps of the four reclassified individual models and the combined model for *Eryngium alpinum* and b) the proportions of surface covered by each of the habitat suitability classes of the combined models for each study species. HHS=high habitat suitability class, MHS=medium habitat suitability class, and LHS=low habitat suitability class.

Model-based sampling

The spatial distribution of the 240 sample points (30 × 8 species; Fig. 1) was driven by model predictions and the species altitudinal ranges. Since most species were characteristically found at low elevations, high elevations were less sampled.

Field-based model evaluation

The evaluation of the models' predictive success, by the results of the 240 fieldwork samples, reveals rather high NPP values for all species, even for those considered common (Table 4). The PPP, however, shows lower values, especially for the rare species. During our field survey, we found several individuals of the rare species that were not recorded in the initial database, but only considered them as new populations if found at 1 km or more from any known presence. With this criterion, five new populations of *Scorzonera laciniata* were discovered from the 10 sites sampled within the HHS class, one from the 10 sites sampled in the MHS class, and eight in total when additionally including two populations founds by chance. The two later populations were not used in any further analysis. No new populations were found for the two other rare species. In the following, we therefore mainly focus the discussion on *S. laciniata*.

Table 4. Positive predictive power (PPP), i.e. proportion of species presences in the high habitat suitability (HHS) class, and negative predictive power (NPP), proportion of species absences in the low habitat suitability (LHS) class, of the combined models. Values are computed from the 30 specific sample points for each species. NA = Not available.

	Species	Negative predictive power (NPP) [%]	Positive predictive power (PPP) [%]
Rare species	*Cypripedium calceolus*	100	NA
	Eryngium alpinum	100	NA
	Scorzonera laciniata	100	31.3
Common species	*Anthyllis vulneraria*	80.6	61.3
	Astrantia major	85.9	71.4
	Briza media	83.3	79.2
	Heracleum sphondilium	72.6	65.9
	Pulsatilla alpina	96.9	53.2

For six out of eight species, the MBS design proved much more efficient than simple random sampling at discovering new populations. The improvement was particularly strong for the rare *S. laciniata* (Fig. 4), for which the probability of new population discovery was of 50%, using the MBS, *vs* 0% by random sampling. When considering a random sample from all our sampled sites located in *S. laciniata*'s HHS class, the success rate was of 30%.

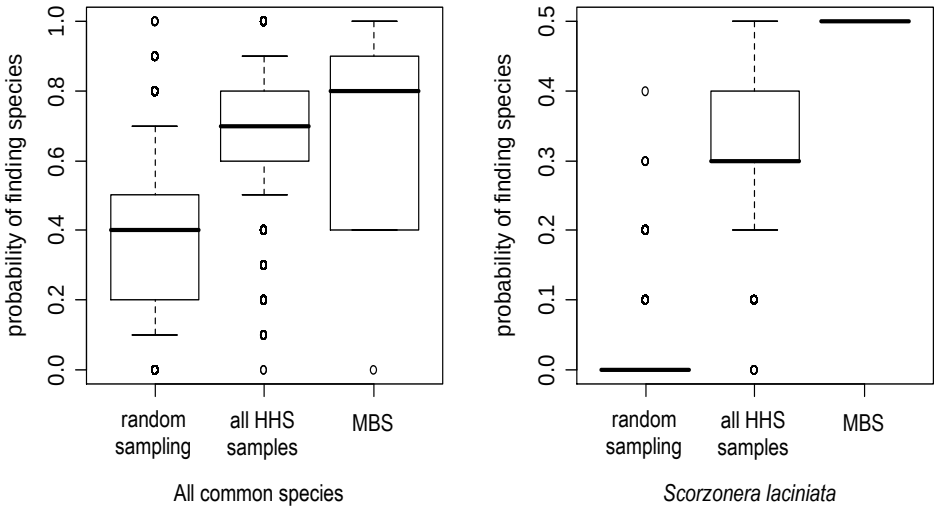


Figure 4. Probability of finding the species when doing random sampling (leftmost box-plot), when looking in all the sampled points located in the HHS class (middle box-plot), and when considering only the ten points selected by MBS (rightmost box-plot). Results are based on data observed during the fieldwork (i.e. 240 sample points), which were pooled for all the common species (left panel). For the rare species (right panel), results are only shown for *Scorzonera laciniata* as no new populations were found for the two others.

Finally, for all species, predictive accuracy of the combined model was higher than that of any individual model (i.e. GAM or ENFA at either resolution), as shown by the weighted Kappa and AUC (Fig. 5 and Table S2 in Supplementary materials).

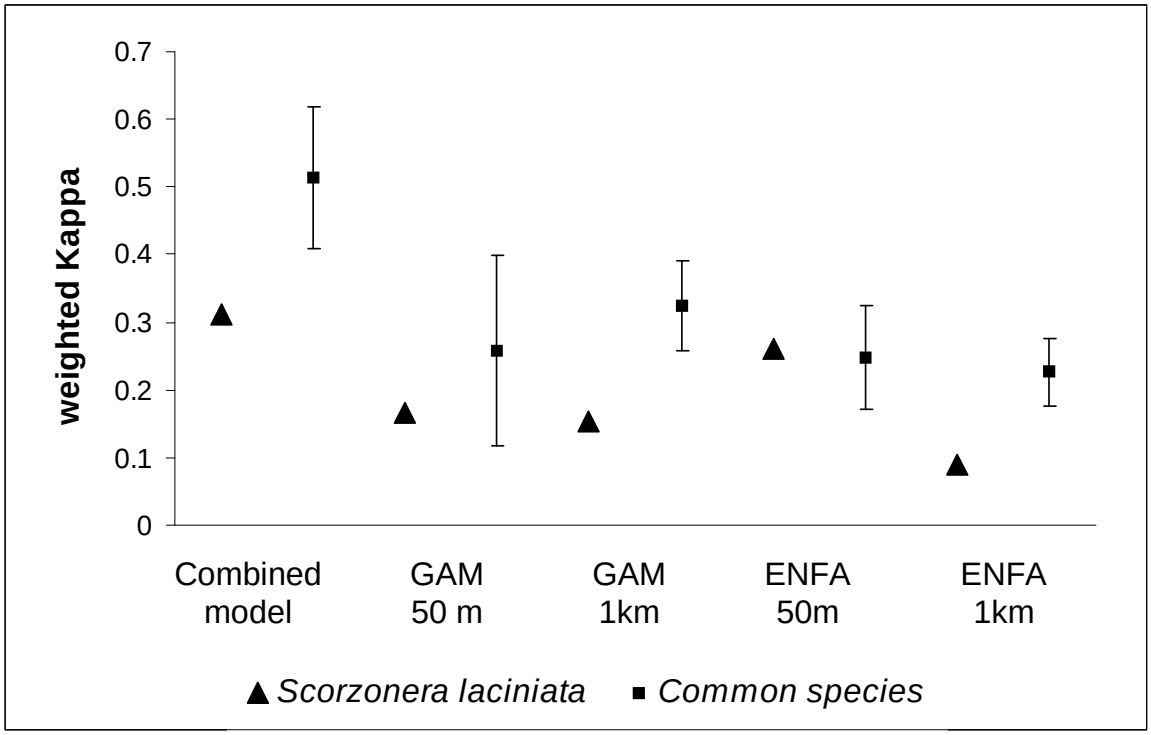


Figure 5. Validation of the models by independent data obtained through fieldwork. For the common species, the weighted Kappa values are pooled, providing mean values and standard deviations. For the rare species, results are only shown for *Scorzonera laciniata*.

Discussion

How efficient is model-based sampling?

In our study, model-based sampling (MBS) proved significantly more efficient than random sampling for finding new populations for six out of our eight target species (Fig. 4). By equally sampling all habitat suitability classes in the field, rather than just the most suitable habitats, we provided an unbiased framework for testing the MBS approach.

For *Scorzonera laciniata*, the MBS approach proved particularly efficient at discovering six new populations out of the ten points sampled in the HHS class. Two additional populations were found by chance within this class, indicating that people may have higher probabilities of finding new populations when looking extensively in the HHS areas.

The positive results obtained for the five common species further evidences the general usefulness of the method to sample any species of interest, not only rare ones. These results corroborate previous findings (Boetsch et al. 2003; Engler et al. 2004; Edwards et al. 2005; Guisan et al. 2006b).

For the two unsuccessful species – *Eryngium alpinum* and *Cypripedium calceolus* – some other parameters may interfere with the modeling results. Among other factors, the

success of finding new populations also depends on the species' detectability (Kéry 2002), the characteristics of the local environments and the sampling effort (Archaux et al. 2006). Given the high degree of rarity and threat of these species, our results could also simply indicate that no additional population exists for *E. alpinum* and *C. calceolus* in the study area.

Overall, our results demonstrate that MBS is particularly efficient to find new populations. Since biodiversity monitoring based on regular sampling, (e.g. Swiss biodiversity monitoring program; http://www.biodiversitymonitoring.ch), result in a very low detection rate of rare species (Bäumler et al. 2005), MBS could make many conservation practices more efficient and cost effective.

How successful is habitat suitability modeling of rare species?

Cross-validation indicated medium to good predictive power for all calibrated models (Table 3). Independent validation (Table 3 and Fig. 5) with fieldwork data revealed more contrasted results. While the sampling of *S. laciniata* and the five common species proved fairly to highly successful, the contrary was observed for *E. alpinum* and *C. calceolus*. For these two species, new populations were not found despite models receiving good evaluation scores (see cv-AUC, Table 3).

Several reasons can explain a mismatch between predictions and observations (Jenkins et al. 2003). First, from a modeling perspective, a model can fail at capturing all the ecological requirements of the species. For example, our models for *E. alpinum* and *C. calceolus* could correctly identify the general trends of the species ecological niche, and thus properly identify their suitable habitats at a regional scale – which returns good cross-validation results, but fail in predicting finer variations of the ecological requirements of the species, i.e. local habitat suitability variations, which causes failures of sampling. This situation could result from incomplete or fuzzy species data (i.e. data precise enough for regional models, but not for local models), or an incomplete set of environmental variables.

Second, a species can be missed in the field, because of its low detectability at the time of the sampling (e.g. too early or too late in the growing season, flowers cut by humans or animals, hidden by local vegetation). Although we tried to minimize these effects as much as possible, one cannot exclude completely this hypothesis. The species could also be present in a biological form that standard sampling method fail to detect, such as seeds, roots or partial vegetative parts when, for instance, plants are eaten by predators or destroyed by human activities.

Third, the absence of a species from HHS areas can be due to fluctuations in population dynamics (e.g. sites too far from source populations, barriers to dispersal, Pulliam 2000; Jenkins et al. 2003), or because of predators or human activities (e.g. flower cutting, trampling, herbicides). When species are not at equilibrium with their environment, habitat suitability models have sometimes difficulties to correctly predict areas of potential occurrences, despite returning good evaluation measures (Cianfrani et al. submitted).

Finally, when all the stable populations of a given species have already been identified, the MBS, as any other sampling method, cannot support the discovery of any new population. However, identifying the hidden factors not accounted for in the model that could potentially be responsible for the species' absence, could help assessing the proximal causes of the species' rarity and discussing the best protection measures to adopt.

How successful are the various modeling approaches?

Models showed various predictive performances (Table 3) and differently ranked the importance of environmental predictors. The resulting differences in spatial projections are visible when comparing reclassified maps of individual habitat suitability models. For instance, at the 50 m resolution, the ENFA model built for *Eryngium alpinum* predicts a larger distribution of the HHS class than the GAM (Fig. 3a), but the opposite trend is observed at the 1 km resolution. Divergences in predictions are illustrated for all species by the models' concordance (Fig. 3b). The models more strongly differ for the common species than for the rare species, probably because rare species have narrower ecological niches, making their specific environmental requirements easier to model (Elith et al. 2006). The ecological plasticity of the common species could make their niches more difficult to model and the differences between the modeling techniques more visible.

Our main result regarding modeling techniques is that, for all species, the model combining the two resolutions and the two modeling techniques is always significantly better than any model considered individually (Fig. 5 and Table 3). This gives support to our initial expectation that combining different resolutions and modeling techniques would allow optimal use of existing data, leading to better models and ultimately improving the ability to detect populations in the field. While our goal was to test a simple method for combining predictions that should be easier to implement for managers, further studies could consider combining more than two modeling techniques and use different methods to combine them (e.g. Marmion et al. in press).

Conclusion

Our study demonstrated the usefulness of the model-based sampling (MBS) approach for enhancing the detection of target species in the field, especially for rare and endangered species. This method could be largely used, for all types of organisms, and in various environments, even in difficult one such as marine ecosystems. We particularly expect the efficiency of MBS to be significantly higher than that of random or regular sampling strategies in difficult fieldwork situations, such as when working with species having low detectability or prevalence, in environments that are difficult to prospect, or across large study areas. Indeed, in these situations, costs associated with fieldwork are usually very high and samplings directed toward sites where chances to find the species are greater would thus be essential. Expert knowledge could usefully complete the modeling results for selecting the best sites when HHS units are large.

We also demonstrated the usefulness of combining modeling techniques (i.e., ensemble forecasting) and resolutions to achieve more robust predictions. One drawback of the MBS methodology is that it requires some initial data on species distribution to produce the first model. However, there are increasing amounts of data from natural history collections (NHMs, Graham et al. 2004) becoming publicly accessible through online databases, such as the Global Biodiversity Information Facility (http://www.gbif.org). Easy-to-use software packages for modeling species distribution (e.g. see Guisan & Thuiller 2005) should also facilitate the implementation of MBS in nature conservation offices. Furthermore, using MBS as fieldwork strategy will not only give insights on model quality by providing

independent evaluation data, but will also help reducing bias in the initial dataset, if all the habitat suitability classes are sampled (Boetsch et al. 2003; Guisan et al. 2006a). Interpretation of the results will additionally help in assessing and understanding potential environmental factors acting on species threats.

We conclude that, from a management perspective, habitat suitability modeling has now reached a stage where it can prove useful to support many conservation practices. We recommend its use for supporting fieldwork and reducing the costs of rare plant management. It could help for various issues such as locating populations, as illustrated in the present study, focusing protection measures on the most valuable habitats where the species is still present (Pawar et al. 2007), or suggesting suitable reintroduction sites (South et al. 2001).

Acknowledgements

We thank the MAVA foundation for financial support and the Swiss Floristic Data Center (CRSF) for providing species data. We thank Pascal Vittoz and Christophe Randin for valuable insights, and all the people who participated in the fieldwork, especially Yannick Delorme, Christian Benetollo, Grégoire Vuissoz and Léa Kissling.

References

Aitken, M., D. W. Roberts, and L. M. Shultz. 2007. Modeling distributions of rare plants in the Great Basin, western North America. West. North Am. Naturalist **67**:26-38.

Akaike, H. 1973. Information theory and extension of the Maximum Likelihood principle. Pages 267-281 in B. N. Petrov, and F. Csaki, editors. Second International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary.

Araújo, M. B., and M. New. 2007. Ensemble forecasting of species distributions. Trends in Ecology and Evolution **22**:42-47.

Archaux, F., F. Gosselin, L. Berges, and R. Chevalier. 2006. Effects of sampling time, species richness and observer on the exhaustiveness of plant censuses. Journal of Vegetation Science **17**:299-306.

Bäumler, B., D. M. Moser, A. Gygax, C. Latour, and N. Wyler. 2005. Fortschritte in der Floristik der Schweizer Flora (Gefässpflanzen). Botanica Helvetica **115**:83-93.

Boetsch, J. R., F. K. Van Manen, and J. D. Clark. 2003. Predicting rare plant occurrence in Great Smoky Mountains National Park, USA. Natural Areas Journal **23**:229-237.

Cianfrani, C., G. Le Lay, A. Hirzel, and A. Loy. submitted. Do habitat suitability models reliably predict the recovery areas of threatened species?

Clot, F., and R. Delarze. 1998. Banque de données phyto des forêts vaudoises. Résultats de l'analyse des relevés du Nord vaudois. Service des forêts, de la faune et de la nature du Canton de Vaud, Lausanne, Switzerland.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement **20**:37-46.

Cohen, J. 1968. Weighted Kappa - Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin **70**:213-220.

Côté, I. M., and J. D. Reynolds. 2002. Predictive ecology to the rescue? Science **298**:1181-1182.

Droz, J. 1994. La végétation de la région de Derborence (Conthey, Chamoson, Valais). Page 239. Institut de Botanique Systématique et de Géobotanique. University of Lausanne, Lausanne.

Edwards, T. C. J., R. Cutler, N. E. Zimmermann, L. Geiser, and J. Alegria. 2005. Model-based stratifications for enhancing the detection of rare ecological events: lichens as a case study. Ecology **86**:1081-1090.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology **41**:263-274.

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.

Gonseth, Y., T. Wohlgemuth, B. Sansonnens, and A. Buttler 2001. Les régions biogéographiques de la Suisse. Explications et division standard. Cahier de l'environnement n° 137. Office fédéral de l'environnement, des fôrets et du paysage., Berne.

Graham, C. H., S. Ferrier, F. Huettman, C. Moritz, and A. T. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology and Evolution **19**:497-503.

Griffin, S. C., B. L. Walker, and M. M. Hart. 2003. Using GIS to guide field surveys for timberline sparrows in northwestern Montana. Northwest Science **77**:54-63.

Guisan, A., O. Broennimann, R. Engler, M. Vust, N. G. Yoccoz, A. Lehmann, and N. E. Zimmermann. 2006a. Using niche-based models to improve the sampling of rare species. Conservation Biology **20**:501-511.

Guisan, A., A. Lehmann, S. Ferrier, M. Austin, J. M. C. Overton, R. Aspinall, and T. Hastie. 2006b. Making better biogeographical predictions of species' distributions. Journal of Applied Ecology **43**:386-392.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters **8**:993-1009.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling **135**:147-186.

Hastie, T., and R. Tibshirani. 1986. Generalized additive models. Statistical Sciences **1**:297-318.

Hirzel, A., and A. Guisan. 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling **157**:331-341.

Hirzel, A., J. Hausser, and N. Perrin. 2006a. Biomapper3.2. Laboratory for Conservation Biology, University of Lausanne, Lausanne.

Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: How to compute habitat- suitability maps without absence data? Ecology **83**:2027-2036.

Hirzel, A. H., and G. Le Lay. 2008. Habitat suitability modeling and niche theory. Journal of Applied Ecology **45**:1372-1381.

Hirzel, A. H., G. Le Lay, V. Helfer, C. Randin, and A. Guisan. 2006b. Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling **199**:142-152.

Jenkins, C. N., R. D. Powell, O. L. Bass, and S. L. Pimm. 2003. Why sparrow distributions do not match model predictions. Animal Conservation **6**:39-46.

Kéry, M. 2002. Inferring the absence of a species - A case study of snakes. Journal of Environmental Management **66**:330-338.

Lehmann, A., J. M. Overton, and J. R. Leathwick. 2002. GRASP: generalized regression analysis and spatial prediction. Ecological Modelling **157**:189-207.

Lodge, D. M., S. Williams, H. J. MacIsaac, K. R. Hayes, B. Leung, S. Reichard, R. N. Mack, P. B. Moyle, M. Smith, D. A. Andow, J. T. Carlton, and A. McMichael. 2006. Biological invasions: Recommendations for US policy and management. Ecological Applications **16**:2035-2054.

Margules, C. R., and R. L. Pressey. 2000. Systematic conservation planning. Nature **405**:243-253.

Marmion, M., M. Luoto, R. K. Heikkinen, and W. Thuiller. in press. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecological modelling.

McKenzie, D., D. W. Peterson, D. L. Peterson, and P. E. Thornton. 2003. Climatic and biophysical controls on conifer species distributions in mountain forests of Washington State, USA. Journal of Biogeography **30**:1093-1108.

Miller, R. I. 1986. Predicting rare plant distribution patterns in the southern Appalachians of the south-eastern U.S.A. Journal of Biogeography **13**:293-311.

Moisen, G. G., and T. S. Frescino. 2002. Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling **157**:209-225.

Pawar, S., M. S. Koo, C. Kelley, M. F. Ahmed, S. Chaudhuri, and S. Sarkay. 2007. Conservation assessment and prioritization of areas in Northeast India: Priorities for amphibians and reptiles. Biological Conservation **136**:346-361.

Pulliam, H. R. 2000. On the relationship between niche and distribution. Ecology Letters **3**:349-361.

Pulliam, H. R., and B. Babbitt. 1997. Science and the protection of endangered species. Science **275**:499-500.

Randin, C. F., T. Dirnbock, S. Dullinger, N. E. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography **33**:1689-1703.

South, A. B., S. P. Rushton, D. W. Macdonald, and R. Fuller. 2001. Reintroduction of the European beaver (Castor fiber) to Norfolk, UK: a preliminary modelling analysis. J. Zool. **254**:473-479.

Swets, J. A. 1988. Measuring the Accuracy of Diagnostic Systems. Science **240**:1285-1293.

Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. Trends in Ecology and Evolution **16**:446-453.

Zimmermann, N. E., and F. Kienast. 1999. Predictive mapping of alpine grasslands in Switzerland: species versus community approach. Journal of Vegetation Science **10**:469-482.

Supplementary Material

Appendix S1: Environmental variable scores

While topo-climatic variables proved overall useful predictors to explain the distribution of the eight species in our study area, their success varied across species, spatial resolution and modeling technique. Moisture index (MIND) was retained in all GAMs. Slope (SLOP), solar radiations (SRAD), vegetation index (NDVI) and potential soil permeability (PERM) were additionally retained in 80% of the GAMs for the common species. SLOP was retained in 70% of the GAMs for the rare species. Contribution of the other variables ranged from 0 to 60%. For the rare species, the main difference between GAM results at the two resolutions is that SRAD was retained in 80% of the GAM at 1 km resolution whereas it was not retained in any of the GAM at 50 m resolution. For the common species, SLOP was retained in 100% of the 1 km GAMs compared to 40% of the 50 m GAMs. In contrast, potential limestone content of soils (CALC) and PERM were both retained in 80% of the 50 m GAMs compared to 60% and 20% of the 1 km GAMs, respectively. For the ENFA models, MIND was also one of the most important variables, for all species and at both resolutions. At the 1 km resolution, CALC was also important for the rare species, complemented by SLOP and the spatial distribution of limestone (GTC3) for the common species. At 50 m, SLOP and NDVI were also important for the common species, whereas the topographic position (TOPO) and MIND were important for rare species.

Table S2.

Predictive performance measured by AUC and Weighted Kappa (WKappa) for the four individual models (two modeling methods and two resolutions) and the combined model for each species. The evaluation is based on presences and absences data gathered during the field work (240 sample points). Bold values indicate best results.

	Species	Validation index	Combined model	GAM		ENFA	
				50 m	1 km	50 m	1 km
Rare species	*Scorzonera laciniata*	WKappa	**0.31**	0.17	0.15	0.26	0.09
		AUC	NA	**0.98**	**0.98**	0.95	0.89
	Anthyllis vulneraria	WKappa	**0.44**	0.03	0.27	0.34	0.27
		AUC	NA	0.54	0.67	**0.72**	0.71
	Astrantia major	WKappa	**0.59**	0.30	0.37	0.30	0.27
		AUC	NA	0.74	**0.79**	0.70	0.71
Common species	*Briza media*	WKappa	**0.64**	0.40	0.40	0.16	0.16
		AUC	NA	0.72	**0.76**	0.60	0.65
	Heracleum sphondylium	WKappa	**0.38**	0.22	0.24	0.18	0.20
		AUC	NA	0.64	**0.67**	0.64	**0.67**
	Pulsatilla alpina	WKappa	**0.51**	0.33	0.35	0.26	0.24
		AUC	NA	**0.87**	0.85	0.81	0.74

Chapter 2

Moving forwards: from static to dynamic species distribution models

2.1 MIGCLIM: Predicting plant distribution and dispersal in a changing climate.

2.2 Seed dispersal distances: a typology based on dispersal modes and plant traits.

2.3 Predicting future distributions of mountain plants under climate change: does dispersal capacity matter?

2.1

MIGCLIM: Predicting plant distribution and dispersal in a changing climate.

Robin Engler[1] and Antoine Guisan[1]

This manuscript has been published in 2009 *Diversity and Distributions*, **15**, 590-601.
DOI: 10.1111/j.1472-4642.2009.00566.x

[1] Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland.

CONTRIBUTION TO THE PROJECT:
I developed the original idea in collaboration with AG. I developed the MIGCLIM cellular automaton, ran all simulations, analyzed the results and wrote the initial complete draft of the manuscript. I also lead the reviewing of the manuscript.

Abstract

Aim: Many studies have forecasted the possible impact of climate change on plant distributions using models based on ecological niche theory, but most of them have ignored dispersal-limitations, assuming dispersal to be either unlimited or null. Depending on the rate of climatic change, the landscape fragmentation and the dispersal capabilities of individual species, these assumptions are likely to prove inaccurate, leading to under- or overestimation of future species distributions and yielding large uncertainty between these two extremes. As a result, the concepts of "potentially suitable" and "potentially colonisable" habitat are expected to differ significantly. To quantify to what extent these two concepts can differ, we developed MIGCLIM, a model simulating plant dispersal under climate change and landscape fragmentation scenarios. MIGCLIM implements various parameters, such as dispersal distance, increase in reproductive potential over time, landscape fragmentation or long distance dispersal.

Location: Western Swiss Alps.

Methods: Using our MIGCLIM model, several simulations were run for two virtual species by varying dispersal distance and other parameters. Each simulation covered the hundred-year period 2001-2100 and three different IPCC-based temperature warming scenarios were considered. Results of dispersal-limited projections were compared to unlimited and no-dispersal projections.

Results: Our simulations indicate that: (i) using realistic parameter values, the future potential distributions generated using MIGCLIM can differ significantly (up to more than 95% difference in colonised surface) from those that ignore dispersal; (ii) this divergence increases under more extreme climate warming scenarios and over longer time periods; (iii) the uncertainty associated with the warming scenario can be as large as the one related to dispersal parameters.

Main conclusions: Accounting for dispersal, even roughly, can importantly reduce uncertainty in projections of species distribution under climate change scenarios.

***Keywords*:** Cellular automaton, climate change, dispersal modelling, dynamic niche-based models, GLM, plant species distribution.

***Abbreviations*:** LDD = long-distance dispersal.

Introduction

Assessing the potential impact of predicted global climate change (+1.8 to +4°K by 2100; IPCC, 2007) on vegetation is a pressing matter. A broad range of projections of future plant species distributions have already been made using niche-based species distribution models (Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005), at large scales (e.g. Bakkenes *et al.*, 2002; Thuiller *et al.*, 2005) and at more local scales (e.g. in alpine landscapes; Gottfried *et al.*, 1999; Dirnbock *et al.*, 2003). Yet, most of these studies have assumed that dispersal is unlimited (or "universal"), where a species is allowed to disperse freely with its future distribution becoming the entire area projected as suitable by the model (Thomas *et al.*, 2004). Because "unlimited dispersal" represents an unrealistic best case scenario, some authors (e.g. Thomas *et al.*, 2004; Thuiller *et al.*, 2005) have also provided a worst case "no dispersal" scenario to give a lower bound to their projections. However, the difference between these extreme projections can yield large uncertainties (e.g. Thuiller, 2004).

While the unlimited dispersal assumption might provide good approximations for plants that are human-dispersed or have high dispersal ability, it likely overestimates the future distribution of many other species because: (i) human-driven habitat fragmentation makes it increasingly difficult for many species to migrate (Pitelka *et al.*, 1997) and (ii) the speed of the expected global warming is predicted to be rapid – one or more orders of magnitude faster than past climate change events (Etterson & Shaw, 2001). Certain species may thus require migration rates much faster than those observed during post-glacial times (Malcolm *et al.*, 2002) and could fail to keep pace with rapid climate change. Compounded with this is the possibility that rapid climate change may not allow many species to evolve the necessary adaptations (Davis & Shaw, 2001; Etterson & Shaw, 2001).

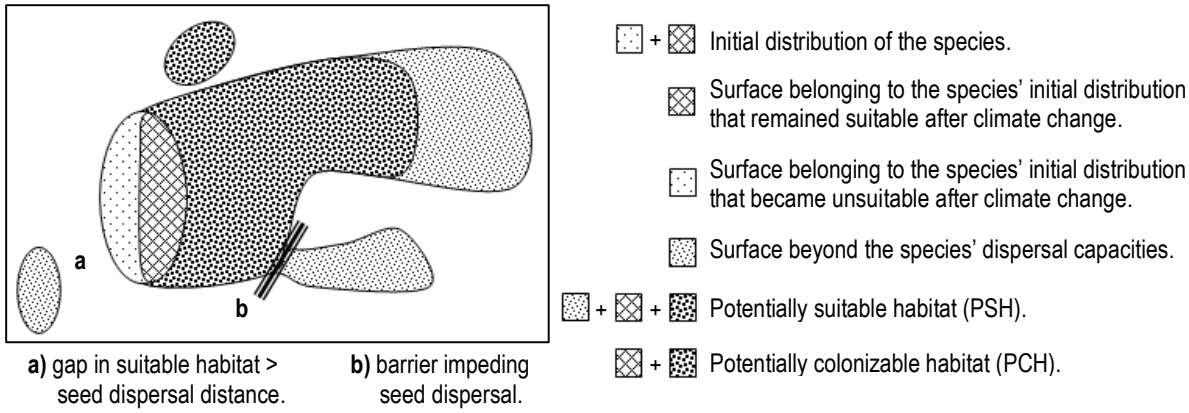


Figure 1. Schematic representation of a species distribution under climate change to explain the concepts of potentially *suitable* and potentially *colonisable* habitat.

Even though the need for including dispersal limitations into models has been raised repeatedly (Pitelka *et al.*, 1997; Woodward & Beerling, 1997; Cain *et al.*, 1998; Davis *et al.*, 1998; Nathan & Muller-Landau, 2000; Ronce, 2001; Araujo & Guisan, 2006; Midgley *et al.*, 2007; Thuiller *et al.*, 2008), so far relatively few niche-based climate change modelling studies have considered it (e.g. Carey, 1996; Ostendorf *et al.*, 2001; Dullinger *et al.*, 2004; Iverson *et*

al., 2004). Accounting or not for dispersal is equivalent to making a distinction between "potentially *suitable* habitat" and "potentially *colonisable* habitat". Potentially suitable habitat is the area a species could occupy given unlimited dispersal ability, whereas potentially colonisable habitat accounts for dispersal constraints such as limited seed dispersal distance, time to reach reproductive maturity, gaps in suitable habitat or barriers to dispersal, e.g. rivers, mountain ranges or dense urban areas (Fig. 1).

Refining the projections from potentially *suitable* to potentially *colonisable* habitat requires the transition from a static to a more dynamic approach of modelling species distributions. One way to achieve this is to use a cellular automaton, which can be briefly described as a grid of cells with values that evolve over time according to rules that are affected by neighbouring cells (Sarkar, 2000). Cellular automata have already been widely used in ecology, for instance to simulate vegetation succession (Yemshanov & Perera, 2002; Scheller *et al.*, 2007), competition between plants (Silvertown *et al.*, 1992), spread of invasive species (Higgins *et al.*, 2000) or plant migration (Collingham *et al.*, 1996; Pearson & Dawson, 2005). Yet, only few studies have employed them to simulate species migrations when predicting future distributions under climate change (e.g. Carey, 1996; Ostendorf *et al.*, 2001; Dullinger *et al.*, 2004), and even fewer have derived projections for many species (e.g. Iverson *et al.*, 2004).

To investigate how dispersal limitations can modify projections of future plant distributions, we developed a model – MIGCLIM – able to simulate plant migration under climate change and landscape fragmentation scenarios. Our MIGCLIM model couples predictive distribution maps, representing a species' habitat suitability as a function of evolving climate, with a cellular automaton that simulates dispersal, colonisation, growth and extinction of the species in the landscape. A number of specific parameters can be defined, such as dispersal distance, barriers to dispersal, landscape fragmentation, stochastic long distance dispersal or increase of reproductive potential over time. Our objective when developing MIGCLIM was to provide a flexible tool that could be used to obtain refined projections of climate change impact for numerous plant species at various spatial scales. Although dispersal models exist based on more advanced population dynamics (e.g. Lischke *et al.*, 2006), they are usually restricted to trees and require advanced knowledge of a species' population dynamics, which is not available for most species. Hence from the perspective of niche-based species distribution models, MIGCLIM represents an important step forward, because it allows more realistic projections of future distributions to be made for many species in a relatively simple way.

Using MIGCLIM, we illustrate how dispersal limitations can affect projections of future plant distributions under climate change compared to unlimited and no dispersal scenarios. We do this by running different simulations for two virtual species under three temperature warming scenarios for the 2005-2100 time period in a study area of the western Swiss Alps. Our study assesses: (i) the possible discrepancies between "potentially suitable" versus "potentially colonisable" habitats; (ii) the factors that may be responsible for these differences.

Methods

The MIGCLIM model

MIGCLIM is a cellular automaton implemented within the ArcGIS® software (ESRI Inc., Redlands, USA). The cells (hereafter pixels) are square and record various values such as pixel occupancy status, habitat suitability, reproductive potential or when it was last colonised. To simulate dispersal under climate change, MIGCLIM requires the following inputs: a map defining the species' initial distribution, maps picturing landscape fragmentation (i.e., barriers to dispersal and permanent unfavourable locations), the species' dispersal parameters and a series of maps indicating how the distribution of potentially suitable habitats evolves as climate changes. Dispersal is simulated through a number of decisions that are taken, for each pixel, during each dispersal event (see also flowchart of Appendix S1 in Supporting Information):

1. Does the target pixel represent a suitable habitat? Is it unoccupied?
2. If point 1 is answered positively, the number n of source pixels within the specified dispersal distance is computed. Source pixels are already occupied pixels that can act as seed sources to colonise a target pixel. Optionally, a barrier layer can be given to prevent dispersal through those pixels being part of the barrier. If a barrier pixel is found between the target and a source pixel, the source pixel is ignored. Barriers can be used e.g. to prevent a strictly grassland species to disperse through forests.
3. If $n > 0$, the target pixel becomes colonised with the combined probability P_{Col} (Eq. 1):

$$P_{Col} = 1 - \prod_{i=1}^{n} (1 - P_{Disp_i} \times P_{Mat_i}) \qquad \text{(Eq. 1)}$$

   Where $P_{Disp\,i}$ is a probability function of the distance between the target pixel and source pixel i and reflects the fact that colonisation probability decreases over distance. $P_{Mat\,i}$ is a probability that is function of the time since the source pixel i became occupied and represents the increase in reproductive potential of source pixel i over time. P_{Mat} can be used to represent time for individuals to reach reproductive maturity and, more globally, the increase of a population's reproductive potential due to an increase in the number of individual plants within a pixel over time. P_{Disp} and P_{Mat} are implemented as discrete functions and can easily be modified to fit any shape of seed dispersal curve and increase of reproductive potential over time.

4. Optionally, long distance dispersal (LDD) and stochastic extinction events can be added to the simulation. LDD events are generated from source pixels with a probability $P_{LDD} \times P_{Mat}$ in a random direction and at a random distance within a user-defined range. If the pixel reached by the long dispersing seed is potentially suitable (satisfying point 1), it becomes colonised. LDD events are not affected by barriers. Stochastic extinctions with probability P_{Ext} can also be defined to simulate random extinction of colonised pixels.
5. Steps 1 to 4 are repeated a number of times (N_{Disp}), typically set so that each repetition corresponds to one year.
6. Pixels that are no longer suitable due to changes in environmental conditions have their values reset to zero. Pixels that become unsuitable are reset only after the dispersion stage occurred (steps 1 to 5), because it is assumed that the change of a habitat from

suitable to unsuitable is not a discrete but a continuous process. Thus, organisms inhabiting a pixel still have the potential to disperse during the step when the pixel turns unsuitable.

7. Steps 1 to 6 are repeated N_{HSmap} times. In each repetition, the habitat suitability is updated to reflect environmental change (e.g. climate change). Simulations without environmental change can be performed by setting $N_{HSmap} = 1$.

Additional parameters available in MIGCLIM include vegetative and seed bank resilience time, post-dispersal survival and/or habitat invasibility, anisotropic dispersal to simulate dominant winds or slope, as well as specific dispersal along certain features, such as roads and rivers. These options were not used in the present study.

Because the smallest modelling unit in the model is a population in a pixel, parameter values must represent values for a whole population, not for a single individual (see also discussion section for more details on model parameterization). All parameters can be modified to best fit the known dispersal characteristics of the modelled species.

Simulations using MIGCLIM

Sensitivity analyses were run for two virtual species in a real landscape (the western Swiss Alps). To mirror ecological reality, the potential distributions under current and future climate conditions of the two virtual species were derived from those of two real species - *Arrhenatherum elatius L.* and *Lolium perenne L.*. Therefore we refer to our virtual species hereafter as ARELA and LOPER. For each species, several simulations were run by varying the dispersal parameters within a range of realistic values (Table 1). These were chosen to represent various plant functional types, and are based on a thorough review of literature (Vittoz & Engler, 2007). Because no accurate data could be found for long distance dispersal (LDD), these were set to 20 times the regular dispersal distance for 5m to 500m simulations and to 10 times the regular dispersal distance for 1000m simulations. This resulted in LDD distances ranging from 100 m to 10 km (Table 1), which corresponds to values commonly found in the literature (Nathan, 2005 and references therein). Two different rates of increase in pixel reproductive potential (P_{Mat}) were also tested; these resembled rates for herbs and trees respectively. In "Herb-type" simulations, the reproductive potential of a pixel followed a sigmoid increase from 1 to 5 years, whereas in "Tree-type", it started at 10 years and reached its maximum at 40 years. Some Tree-type simulations were not run for low dispersal distances because these were not considered realistic for trees. The spatial resolution (i.e., pixel size) at which the simulations were run was of either 25 m, 12.5 m or 5 m, depending upon the maximum dispersal distance of the simulation (see Table 1).

Table 1. Dispersal parameters of the different ARELA and LOPER simulations. "H" and "T" in the simulation name stand for "Herb-type" and "Tree-type", and refer to the increase in pixel reproductive potential over time that reflects either a herb-like or tree-like life cycle.

simulation name	5m H	25m H	100m H	500m H	1000m H	100m T	500m T	1000m T
Pixel size	5 m		12 .5 m	25 m		12.5 m	25 m	
disp. event frequ. *	5 years	1 year						
dispersal dist. (*DispDist*)	5 m	25 m	100 m	500 m	1'000 m	100 m	500 m	1'000 m
max LDD dist.	100 m	500 m	2'000 m	10'000 m	10'000 m	2'000 m	10'000 m	10'000 m
LDD frequ. **	0.002	0.0004	0.0025	0.01	0.01	0.0025	0.01	0.01
pix. reprod. pot. ***	Sigmoid increase from 1 to 5 years					Sigmoid increase from 11 to 40 years		
Barriers	Yes (forests)					No		
Filter	Urban areas and lakes							
Number of pixels	30'000'000		4'800'000	1'200'000		4'800'000	1'200'000	

* Dispersal event frequency indicates the time between two successive dispersal events.

** Variations in LDD event frequencies between simulations reflect the corrections applied for cell size and dispersal event frequency to maintain the 0.01 frequency assigned to 25 m pixels constant over all categories.

*** Pixel reproductive potential. Indicates how the probability of a colonised pixel to be a source pixel increases over time.

The decrease in colonisation probability with increasing distance from a source cell (P_{Disp}) was based on a negative exponential seed dispersal probability distribution function (Eq. 2; Ward *et al.*, 2004; Scheller *et al.*, 2007):

$$P_{seed}(x) = e^{(x-pixelsize)\cdot\left(\frac{\ln(1-k)}{DispDist}\right)} - e^{x\cdot\left(\frac{\ln(1-k)}{DispDist}\right)} \qquad \text{(Eq. 2)}$$

which can be simplified into the more conventional simple negative exponential form (Eq. 3):

$$P_{seed}(x) = \left(\left(1-k\right)^{-\frac{pixelsize}{DispDist}} - 1\right)\cdot e^{x\cdot\left(\frac{\ln(1-k)}{DispDist}\right)} \qquad \text{(Eq. 3)}$$

where P_{seed} is the probability of a seed reaching distance $x \geq pixelsize$, *pixelsize* is the one-dimensional size of a pixel, and *DispDist* is the dispersal distance reached by the proportion k of the seeds. Here we set k = 0.99, *DispDist* thus represents the regular dispersal distance for seeds, i.e., LDD events excluded (these were here modelled as a separate process). Since the surface composed of pixels located at distance j from a source cell increases with distance from that source cell, the probability of a pixel to receive a seed is computed as (Eq. 4):

$$P_{seed}(pixel_j) = {P_{seed}(x)}\Big/{Surface_j} \qquad \text{(Eq. 4)}$$

where $Surface_j$ is the surface covered by all pixels belonging to a same distance class. Assuming that the distribution of successful seeds (i.e., seeds leading a pixel to become colonised) is proportional to the overall distribution of seeds (P_{seed}), P_{Disp} is computed as (Eq. 5):

$$P_{Disp}(pixel_j) = 1 - (1 - P_{seed}(pixel_j))^{Successful\,Seeds} \quad \text{(Eq. 5)}$$

where P_{Disp} is the probability of colonisation for a target pixel with distance *j* from a source pixel and *Successful Seeds* the number of successful seeds produced by a fully mature source pixel. Since the value of *Successful Seeds* was unknown for our virtual species (as is the case for most species), it was set so that P_{Disp} = 0.99 for a pixel immediately adjacent to a source pixel at 25 meters resolution. Using this conservative (i.e., optimistic) calibration method led to average spread rates, for fully mature pixels, between 45% and 85% of *DispDist*, depending on the number of source pixels in the neighbourhood (tests run on homogenous landscapes). When running simulations at smaller pixels sizes (i.e., 12.5 m and 5 m), the values of P_{Disp} were recomputed in order to ensure that the production of *Successful Seeds* per surface unit remained constant. In other words, the number of *Successful Seeds* was always proportional to pixel size, 5 m and 12.5 m pixel producing respectively 25 and 4 times less successful seeds than 25 m pixels. This ensured that the species had always the same seed production per surface unit and thus that their spread rate was independent of the cell size at which simulations were run. Details of all dispersal kernels used in the different simulations are given in appendix S2.

A negative exponential kernel shape as used here for P_{Seed} (Eq. 3) is common for seed dispersal (Willson, 1993), but many alternatives exist (e.g. Clark *et al.*, 1999; Nathan & Muller-Landau, 2000; Greene *et al.*, 2004) that could be used in MIGCLIM. While negative exponential kernels were found to reasonably approximate different seed dispersal types in previous studies (Bleher *et al.*, 2002; Bischoff, 2005; Clark *et al.*, 2005), one shortcoming is their underestimation of LDD events (Bullock & Clarke, 2000; Nathan *et al.*, 2008). In our simulations, however, LDD events were modelled as an additional, separate, process (this is an option in MIGCLIM). Hence, the negative exponential curve remains a reasonable approximation for regular dispersal (i.e., LDD excluded). Furthermore, while MIGCLIM can accommodate any shape of dispersal kernel, it can only be as refined as the pixel resolution allows. For instance, at a 5 m pixel resolution, a dispersal distance of 20 m can only be approximated by a dispersal kernel with 4 knots. Increasing pixel resolution, as done here for simulation with shorter dispersal distances (Table 1), would theoretically allow infinite refinement of the dispersal kernel. In practice however, computing power limits the minimal pixel size that can be used for a given study area. In our case, 5 m resolution was the smallest we could use, resulting in more than 30 million pixels for our study area.

For each species, all simulations listed in Table 1 were run for three temperature warming scenarios (3 levels; Appendix S3) and both with and without LDD events (2 levels). Each of these runs was replicated 30 times. The temperature increase scenarios represent the maximum ("max", + 5.8°K), intermediate ("med", + 3.6 °K) and minimum ("min", + 1.4 °K) warming scenarios for 2100 of the intergovernmental panel on climate change third assessment report (Houghton *et al.*, 2001). The change in habitat suitability was modelled every fifth year from 2005 to 2100. The initial distribution of the species was set to their potentially suitable habitat under current climate conditions. Dispersal was simulated on a yearly basis, except for the 5m Herb-type simulations where dispersal was simulated every fifth year and the dispersal distance multiplied by five (this assumes a conservative "best case" approximation where spread rate = generation time × dispersal distance). This was

done because current computing power did not allow for a finer grid size than 5 meters for our entire study area. No- and unlimited-dispersal simulations were also carried out for each climate change scenario to provide lower and upper bounds to the projections.

Study area and virtual species

The study area is a 700 km^2 subset of the western Swiss Alps (Fig. 2), with altitudes ranging from 400 to 3,200 meters a.s.l. *Arrhenatherum elatius* and *Lolium perenne*, the two grass species from which we derived the virtual species' distributions, are dominant in the study area and therefore are good indicators of possible change in the vegetation. ARELA illustrates the case of a species whose potentially suitable habitats are purely expanding (i.e., no habitat becomes unsuitable as climate changes) while LOPER illustrates the case where suitable habitat is shifted upward in elevation.

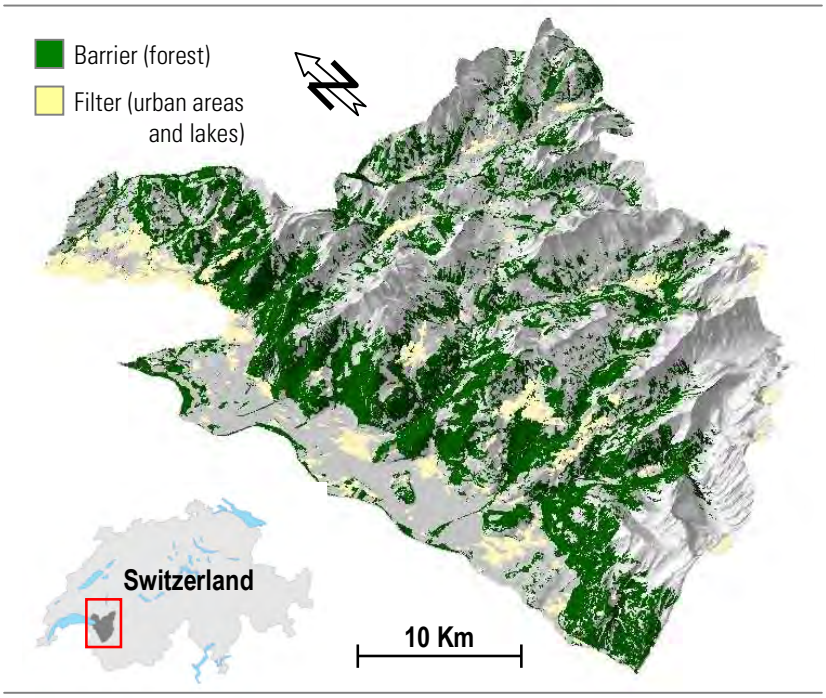


Figure 2. View of the Swiss western Alps study area showing forests (used as a "barrier" feature in "Herb-type" simulations) as well as urban areas and lakes (used as "filter" features in all simulations).

Habitat suitability modelling

Generalized linear models (GLM; McCullagh & Nelder, 1989) were used to statistically relate presence-absence data to a set of topo-climatic environmental variables. The species distribution data originated from a dataset of 550 vegetation plots, sampled in a random-stratified way (Randin *et al.*, 2006). The set of initial environmental variables is given in Appendix S4. After backward stepwise selection, the final variables retained in the models

were: mean annual temperature, mean annual daily solar radiation (this accounts for topographic overshadowing, slope, and aspect values), and slope. The calibrated GLMs were then projected for current climatic conditions (1961-1990 averages) and for every five year period from 2005 to 2100 to produce probabilistic habitat suitability predictions, which were reclassified into binary presence/absence maps (1 or 0) using a optimized- Kappa threshold (Engler *et al.*, 2004).

Environmental variables were available at a spatial resolution of 25 meters, therefore also defining the resolution of the habitat suitability maps. To enable finer dispersal kernel modelling for the simulations with shorter dispersal distances, the resolution of habitat suitability maps was increased (when needed; see Table 1) to 12.5 or 5 meters by dividing the original 25 meter pixels into 4 or 25 smaller pixels.

Results

All values presented are means of 30 repetitions. Standard deviations are not shown as they never exceeded 0.5% and 1.6% of the mean values obtained for the simulations with and without LDD. The initial distribution and an example of future distribution by 2100 of ARELA and LOPER are given in Appendix S5.

Comparing the projected distributions for 2100 of ARELA and LOPER to their initial distribution reveals that ARELA always increases its distribution as a result of climate change (Fig. 3A, all values $\geq$ 100%), whereas LOPER either loses or gains distribution surface, depending on the scenarios' dispersal distance (Fig. 3A: simulations with short dispersal distances have values < 100%, those with longer dispersal distances values > 100%). Comparing dispersal-limited to unlimited dispersal projections (Fig 3B) by 2100 allows assessing the over-estimation made when using the unlimited dispersal modelling approach. This over-estimation varies from less than 20% when using dispersal distances of 500 m or more, up to about 75% (ARELA) or 95% (LOPER) for 5m Herb-type, 25m Herb-type and 100m Tree-type simulations under maximum climate change.

The impact of dispersal constraints and warming scenarios on distribution projections is modulated by the interaction between change in habitat suitability and landscape configuration. As a result, the importance of the impact of a given climate change scenario is not always the same for each dispersal distance scenario. For instance, the difference in surface between dispersal-limited and unlimited dispersal projections is usually more important under the maximum than under the medium climate change scenario, but exceptions exist (e.g. 500m Herb-type simulations, Fig. 3B).

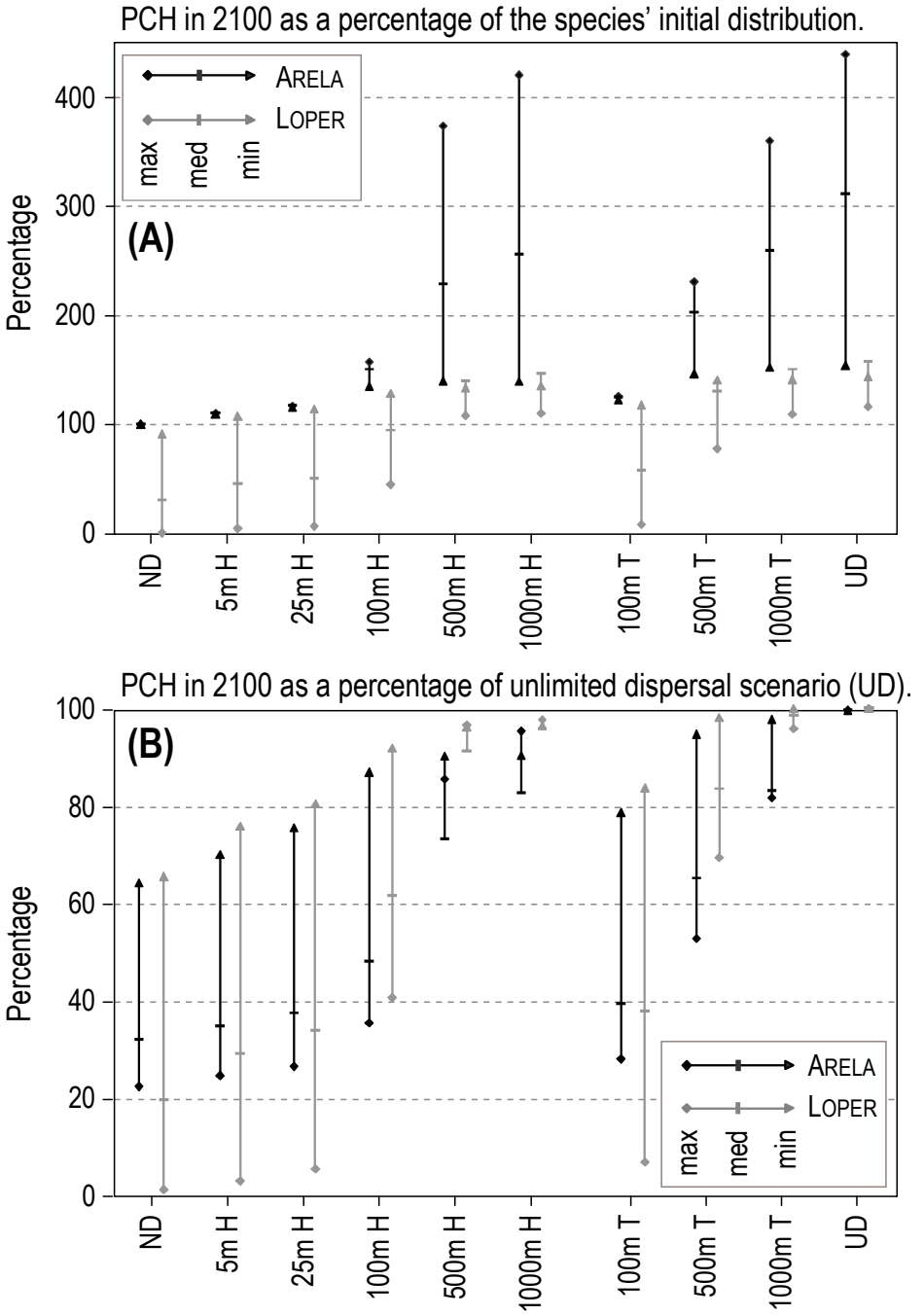


Figure 3. Potentially colonisable habitat (PCH) by 2100 for Arela and Loper simulations without LDD events for the three temperature warming scenarios (max, med, min). 5A) shows the PCH expressed as a percentage of the area initially occupied by a species. 5B) shows the PCH as a percentage of the unlimited dispersal scenario, which equals the total potentially suitable habitat. ND = no dispersal, UD = unlimited dispersal, simulation names refer to those given in Table 1.

Comparing each simulation with LDD to its counterpart without LDD shows that the increase in projected distribution due to LDD remains on average below 50% (i.e., values smaller than 150% in Fig. 4), but can occasionally reach values far greater than 100% (e.g. 360% for LOPER 100m Tree-type simulation; Fig. 4).

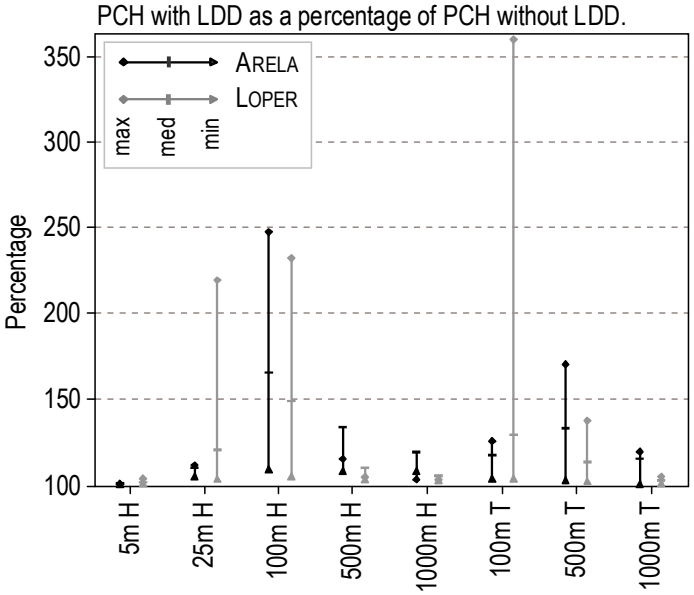


Figure 4. Relative increase of potentially colonisable habitat (PCH) by 2100 for Arela and Loper simulations when introducing LDD events into the simulations for the three temperature warming scenarios (max, med, min). Names of simulations refer to those given in Table 1.

The difference between projections obtained using the different dispersal scenarios is not constant and can increase considerably over time. This is illustrated for LOPER under maximum warming scenario (Fig. 5): while the difference in distribution between the different dispersal scenarios does not exceed 20% of the species' initial distribution in 2020, it becomes larger than 100% by 2100. The abrupt change in projected distributions around 2040 results from the geographical configuration of the study area, with the large flat area of the Rhone valley (western part of the study area, Fig. 2) becoming unsuitable for LOPER around 2040.

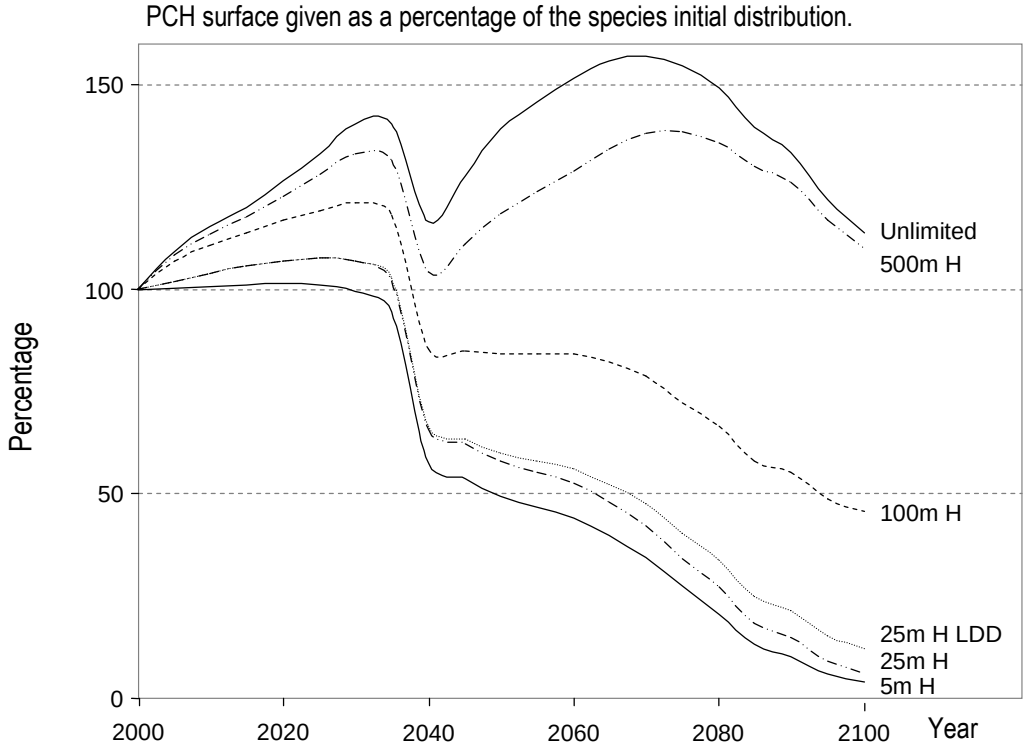


Figure 5. Evolution of the potentially colonisable habitat (PCH) over time for Loper Herb-type simulations under maximum temperature warming scenario. PCH surface is given as a percentage of the species' initial distribution. Names of simulations refer to those given in Table 1.

Discussion

Impact of dispersal limitations on future distribution projections

Results of our simulations illustrate that depending on the warming scenario and the dispersal parameter values, taking dispersal into account can lead to strikingly different predictions from those obtained when dispersal is assumed to be either unlimited or null. For instance, in 5m Herb-type and 100m Tree-type simulations under maximum warming scenario (+ 5.8 °K by 2100), the potentially colonisable habitat predicted for 2100 is about 4 times smaller than the potentially suitable habitat for ARELA, and more than 15 times smaller for LOPER. In other words, predictions made by considering dispersal as unlimited would result in a 4-fold or more than 15-fold over-estimation of the species' future distribution. For the unlimited dispersal scenario to be a good approximation (less than 5% difference between potentially *suitable* and potentially *colonisable* habitat), both species would need to have a regular seed dispersal distance (i.e., LDD excluded) reaching at least 1000 m, or sometimes 500 m if implementing long distance dispersal. Most species in our study area likely have regular dispersal distances much smaller than 1000 m (Vittoz & Engler, 2007), and are therefore likely to have future distributions significantly smaller than predicted by unlimited dispersal scenarios.

Not surprisingly, results obtained under the maximum warming scenario lead to the highest discrepancy between dispersal-restricted and unlimited dispersal projections, with surface differences up to about 75% for ARELA and 95% for LOPER. Yet, even under the medium (+3.6°K) or minimum (+1.4°K) warming scenarios, this discrepancy still reached about 60-70% (medium scenario) and 20-30% (minimum scenario) for both ARELA and LOPER (5m Herb-type or 100m Tree-type simulations, Fig. 3B).

Because for most species no accurate data is available on dispersal distances, only rough estimates can be obtained or derived from the literature, by considering species dispersal characteristics (dispersal vector, seed weight, etc.; Vittoz & Engler, 2007). *Arrhenatherum elatius* and *Lolium perenne*, the two species used to derive the virtual species (ARELA and LOPER), are mainly anemochorous (Müller-Schneider, 1986), suggesting a seed dispersal distance (LDD excluded) of about 20 m (Vittoz & Engler, 2007). This estimate is consistent with dispersal distances measured for other grass species (e.g. 20 m for *Bromus sterilis*; Howard *et al.*, 1991) and is also within the range of distances generally found for non-tree species (10-80 m; Bullock & Clarke, 2000, and references therein). Hence, the 25m and 100m Herb-type simulations are likely to cover a realistic range of dispersal distances for these species and provide a reasonable estimate of discrepancies between projections accounting for dispersal restrictions and those considering dispersal as unlimited or null.

Our results also illustrate how using rough estimates (e.g. dispersal distance between 20 and 100 m, as discussed above), rather than the no-dispersal and unlimited dispersal bounds can strongly reduce uncertainty in projections, even when the exact dispersal distance of the species remains unknown. When considering the maximum warming scenario for ARELA, the ratio in projected distribution surface between unlimited dispersal and no-dispersal is 4.4 (meaning that the surface predicted by unlimited dispersal is 4.4 times larger than that of no-dispersal), but is reduced to 1.3 between 25m and 100m simulations. For LOPER, this ratio is reduced from 71.2 to 7.5. Comparable changes, though of a smaller magnitude, are observed for the medium (3.1 to 1.3 for ARELA and 5.1 to 1.8 for LOPER) and minimum (1.5 to 1.1 for both ARELA and LOPER) warming scenarios.

Effect of long distance dispersal (LDD)

Introducing LDD events in our simulations led to various relative rates of increase in projected distributions (Fig. 4), ranging from no or very little increase (e.g. 5m and 1000m simulations) up to 360% (LOPER 100m Tree-type simulations under maximum warming). While the maximum LDD distance is obviously an important factor determining whether including LDD will have a large effect on projections (e.g. 5m simulations show little increase in potentially colonisable habitat because LDD distance is too short), another aspect is the availability of suitable but unoccupied pixels in the study area. For instance, the 1000m simulations show little increase in potentially colonisable habitat when adding LDD because they already fill almost their entire suitable habitat without LDD. These empty suitable sites can either result from: (i) the species not migrating fast enough to keep up with climate change (explaining why the increase in distribution when introducing LDD is usually highest in the maximum warming scenario); or (ii) the presence of barriers or gaps in suitable habitat that impedes the species from occupying all its potentially suitable habitats (Fig. 1). Therefore, although LDD can already have an important effect on projections made at a local scale, its impact should be even greater at larger, e.g. continental, scales. Our results also illustrate how anticipating the impact of dispersal parameters such as LDD can be difficult, further supporting the need for tools such as MIGCLIM for running sensitivity analyses.

Importance of climate change scenario

Along with dispersal distance, the intensity of the warming scenario proved to be a key factor in determining the level of uncertainty in future projections. In our simulations, variation in predicted future distributions always increased under more severe warming scenarios (Fig. 3 and 4). The uncertainty in projections also increased over time (Fig. 5). As a result, ignoring dispersal or using incorrect dispersal parameters can lead to larger errors when simulations are made for higher temperature warming scenarios or over longer time periods. The variation in projected future distribution related to the temperature warming scenarios is also strongly dependent on the dispersal parameter values: it can be as important as – or even greater than – the variation due to dispersal parameters (e.g. Fig. 3, ARELA 500m or 1000m simulations), but it can also be very small (e.g. Fig. 3, ARELA 5m or 25m simulations). For instance, even if the ARELA seed dispersal distance was identified to be precisely 500 m, a large amount of uncertainty would remain associated with the projections because of the warming scenarios. On the other hand, if one could show that this species only disperses seeds at 5 or 25 m, more accurate predictions could be achieved even though the climate change scenarios remain uncertain.

Generalizing our results

Results of projections for our two virtual species for 2100 showed that differences between potentially *suitable* and potentially *colonisable* habitat could vary from less than 5% to more than 95%, depending on the dispersal distance and warming scenario that were used. The impact of a particular parameter on projections is dependent on: (i) the geographical configuration of the study area, (ii) the initial distribution of the species, and (iii) how the distribution of the potentially suitable habitats evolves as environmental conditions change. For these reasons, caution is recommended before generalizing our results to other areas or to other plant species. How frequent and important the overestimations are remains to be determined, but predictions for slow-dispersing species or made in fragmented landscapes should benefit from constraining projections through dispersal limitations. Furthermore,

because our study area is mountainous, the distances required to keep track of suitable climate conditions are likely much shorter than in a flatter landscape (i.e., less Euclidian distance separates current and forecasted future distribution ranges due to the generally steeper climatic gradient found in mountains, e.g. Midgley *et al.*, 2003). Therefore, the differences found here between potentially *suitable* and potentially *colonisable* habitat (Fig. 3) may well constitute a conservative baseline, with greater differences to be expected when applying MIGCLIM to flatter and larger areas (as found e.g. by Midgley *et al.* (2006) on Proteaceae of the Cape Floristic Region).

Calibration, current limitations and perspectives of MIGCLIM

The higher complexity of dispersal models can also be a potential weakness, as they require a deeper knowledge of the modelled organism and its dispersal characteristics, for which data are often difficult to obtain (see Higgins *et al.*, 2003 for a review). For instance, although methods to estimate LDD events exist (Nathan *et al.*, 2003) and significant improvements have been made in LDD modelling (e.g. Takenberg, 2003; Soons *et al.*, 2004; Nathan *et al.*, 2005), accurately integrating them into spatially explicit models of species distribution and dispersal remains challenging. In fact, due to the inherent unpredictability of LDD, accurate predictions might not even be possible (Clark *et al.*, 2003).

In order to achieve accurate forecasts, the calibration of dispersal parameters must be made with care, especially when working with more extreme climate change scenarios and over long time periods. Ideally, this choice should be based on experimental or historical data (e.g. Iverson *et al.*, 1999), but in practice these are scarce. Nonetheless, our results demonstrate that rough estimates can already significantly reduce uncertainty when compared to the large interval defined by the unlimited and no dispersal scenarios. Below are some points of guidance for calibrating MIGCLIM's parameter:

Calibration of P_{Disp} (probability of a pixel to be colonised given its distance to a source pixel): While accurately fitting a dispersal kernel can be difficult, estimating *DispDist* (i.e., the maximum regular dispersal distance, LDD excluded) is usually possible (see e.g. the methodology proposed in Vittoz & Engler, 2007). In the simplest case, P_{Disp} can be set to 1 for all distances within *DispDist*. While this is a somewhat simplistic option, as it ignores the fact that P_{Disp} decreases with increasing distance from a source pixel, this setting can still be justified as being an optimistic best case approach. The alternative requires defining a seed dispersal kernel (P_{Seed}) and a link function relating P_{Seed} to P_{Disp}. Here we set the link function by assuming that the dispersal of successful seeds follows the same distribution than the dispersal of any seed (this might not be fully correct due to possible density-dependent seed mortality), and calibrated it by setting Eq. 5 equal 0.99 for pixels immediately adjacent to a source cell at 25 m pixel resolution. Although this calibration already leads to significant dispersal limitations, it remains an approximation and might produce overly optimistic dispersal rates. This calibration approach could nevertheless be improved if more data were available: for instance knowing the number of seeds produced per surface unit and the rate with which these seeds survive to become mature individuals would allow computing more accurate values of P_{Disp} using Eq. 5. Another approach for calibrating the link between P_{Seed} and P_{Disp} would be to choose it so that the average spread rate corresponds to historical (e.g. as in Iverson *et al.,* 1999) or recent (e.g. Takahashi & Kamitani, 2004) measures for the species or a species with similar dispersal strategy. Finally, using more complex models (e.g. Pearson & Dawson, 2005) could also be a possible way of exploring and improving the link between P_{Seed} and P_{Disp}.

Calibration of LDD (Long Distance Dispersal): for computational efficiency reasons, we modelled LDD events as a separate process from regular dispersal. Ideally however, and this remains possible in MIGCLIM, both regular and long distance dispersal should be modelled in a single kernel. In such case, the choice of the dispersal kernel becomes extremely important: choosing a fat-tailed kernel will result in several magnitudes more LDD events than a negative exponential kernel (Nathan *et al.*, 2008).

Calibration of P_{Mat} (probability of a pixel to act as source pixel given the time since it became colonised): As illustrated by our results (Fig. 3, comparing Tree-type and Herb-type simulation results), P_{Mat} has an important effect on the projections and can significantly slow down a species' progression. One simple way to calibrate this parameter is matching it to the values of an individual plant, setting it equal 0 before an individual starts producing seeds and equal 1 from then on. Again, this is not fully realistic but represents a best case approach that can already greatly impact predictions, especially for organisms such as trees that need many years before producing seeds. If more detailed knowledge is available, this information can obviously be reflected in P_{Mat}.

Initial species distribution: In our simulations, the initial distributions of ARELA and LOPER were set to their entire potentially suitable habitat under current climatic conditions. While it would be preferable to use the true distribution of a species as starting point, this information is rarely known. A case where the initial distribution might be pretty well known is for invasive species: MIGCLIM could thus use the area of introduction of an invasive species to model its potential spread through the landscape.

Choice of pixel resolution: One obvious point is that pixel size should be much smaller than *DispDist* (here we always used a pixel size at least 5 times smaller than *DispDist*). Another issue that should be borne in mind is that the calibration of P_{Disp} (Eq. 5) is dependent on the pixel size at which a simulation is run (i.e., for a same species, larger pixels should produce more seeds than smaller ones). In this study we developed kernels of P_{Disp} for 12.5 m and 5 m pixel sizes so that the density of produced seeds per surface unit remained the same than for 25 m pixel sizes. This ensured that our species had a spread rate that was independent of the pixel size. Another possibility to keep constant spread rate across different pixel sizes is to use an optional MIGCLIM parameter that allows correcting the values of P_{Disp} when used with a different pixel size than the one it was calibrated for (see Appendix S6 for more details on this correction factor).

Habitat suitability: In our simulations, habitat suitability was treated as a binary variable, pixels being either suitable or not. In reality, habitat suitability is obviously a more continuous variable. This is an important consideration because habitat suitability does, in our model, implicitly represent post-dispersal seed-to-adult survival probability, which is a critical, yet often overlooked, parameter in species dispersal (Nathan, 2006). Using continuous rather than binary habitat suitability values is available as an optional parameter in the MIGCLIM model. The same parameter can also be used to integrate a measure of habitat invasibility (e.g. Dullinger *et al.*, 2004).

The objective while developing MIGCLIM was to provide a flexible tool that can be used to obtain refined projections for many species, for which information available on dispersal characteristics is basic or can only be roughly estimated. Therefore the model must necessarily sacrifice some accuracy for generality and cannot include as many parameters as other models developed for a single species (e.g. Dullinger *et al.*, 2004) or for addressing more theoretical insights (e.g. Pearson and Dawson, 2005). In its current state, MIGCLIM does not include explicit population dynamic parameters, such as number of individuals per pixel or

number of seeds per individual. However, calibrating such parameters requires more complex population dynamic data, which are simply not available for most species.

Conclusion

This study illustrates to which extent "potentially suitable habitat" and "potentially colonisable habitat" can considerably differ when addressing the issue of plant distributions under climate change. Our results further support the view of dispersal being a key factor in this context (Pitelka *et al.*, 1997; Woodward & Beerling, 1997; Davis *et al.*, 1998; Cain *et al.*, 2000; Ronce, 2001).

A main limitation of dispersal models remains their requirement for deeper ecological knowledge in order to calibrate them for a given species. Nonetheless, approximations of these parameters can be obtained (e.g. Vittoz & Engler, 2007), and used to achieve refined projections for a large number of species. Additionally, as illustrated here, sensitivity analyses performed by varying dispersal parameters along a range of realistic values allows to generate complementary uncertainty estimates around existing projections (e.g. Thuiller, 2004; Thuiller *et al.*, 2005).

MIGCLIM is available at www.unil.ch/ecospat.

Supporting Information

Appendix S1. Flow-chart view of the MIGCLIM model.

Appendix S2. Detail of P_{Disp} and P_{Mat} values used in the different simulations.

Appendix S3. IPCC-based temperature increase scenarios used in this study.

Appendix S4. Initial set of environmental variables used as input for the GLM stepwise selection procedure.

Appendix S5. Initial distribution and examples of predictions obtained for year 2100 for ARELA and LOPER.

Appendix S6. Explanation and justification of the optional pixel-size correction factor.

Acknowledgements

The study was supported by the Swiss National Science Foundation (grant nr 110000 and NCCR Plant Survival), the European Commission (FP6 Projects MACIS and ECOCHANGE) and the Herbette foundation. Our warmest thanks go to Niklaus E. Zimmermann, Stephan Dullinger, Wilfried Thuiller, Ran Nathan and three anonymous reviewers for their constructive comments and suggestions on earlier versions of the manuscript, Christophe Randin for help with GLM and environmental data, Roxanne Milleret for early analyses, Alexandre Hirzel for advice on programming, Thomas Czáka for GIS assistance and Cecilia Jordan and Dorothea Pio for linguistic revision.

References

Araujo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677-1688.

Bakkenes, M., Alkemade, J.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.

Bischoff, A. (2005) Analysis of weed dispersal to predict chances of re-colonisation. *Agriculture, Ecosystems and Environment*, **106**, 377–387.

Bleher, B., Oberrath, R. & Böhning-Gaese, K. (2002) Seed dispersal, breeding system, tree density and the spatial pattern of trees – a simulation approach. *Basic and Applied Ecology*, **3**, 115–123.

Bullock, J.M. & Clarke, R.T. (2000) Long distance seed dispersal by wind: measuring and modelling the tail of the curve. *Oecologia*, **124**, 506-524.

Cain, M.L., Brook, G.M. & Strand, A.E. (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany*, **89**, 1217-1227.

Cain, M.L., Damman, H. & Muir, A. (1998) Seed Dispersal and the Holocene Migration of Woodland Herbs. *Ecological Monographs*, **68**, 325-347.

Carey, P.D. (1996) DISPERSE: A cellular automaton for predicting the distribution of species in a changed climate. *Global Ecology and Biogeography Letters*, **5**, 217-226.

Clark, C.J., Poulsen, J.R., Bolker, M., Connor, E.F. & Parker, V.T. (2005) Comparative seed shadows of bird-, monkey-, and wind-dispersed trees. *Ecology*, **86**, 2684–2694.

Clark, J.S., Lewis, M., McLachlan, J.S. & HilleRisLambers, J. (2003) Estimation population spread: what can we forecast and how well? *Ecology*, **84**, 1979-1988.

Clark, J.S., Silman, M., Kern, R., Macklin, E. & Islamber, J.H. (1999) Seed Dispersal Near and Far: Patterns across temperate and tropical forests. *Ecology*, **80**, 1475-1494.

Collingham, Y.C., Hill, M. & Huntley, B. (1996) The migration of sessile organisms: a simulation model with measurable parameters. *Journal of Vegetation Science*, **7**, 831-846.

Davis, A.J., Jenkinson, L.S., Lawton, J.H., Shorrocks, B. & Wood, S. (1998) Making mistakes when predicting shifts in species range in response to global warming. *Nature*, **391**, 783-786.

Davis, M.B. & Shaw, R.G. (2001) Range Shifts and Adaptive Responses to Quaternary Climate Change. *Science*, **292**, 673-679.

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.

Dullinger, S., Dirnbock, T. & Grabherr, G. (2004) Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. *Journal of Ecology*, **92**, 241-252.

Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41**, 263-274.

Etterson, J.R. & Shaw, R.G. (2001) Constraint to Adaptive Evolution in Response to Global Warming. *Science*, **294**, 151-154.

Gottfried, M., Pauli, H., Reitter, K. & Grabherr, G. (1999) A fine-scaled predictive model for changes in species distribution patterns of high mountain plants induced by climate warming. *Diversity and Distributions*, **5**, 241-251.

Greene, D.F., Canham, C.D., Coates, K.D. & Lepage, P.T. (2004) An evaluation of alternative dispersal functions for trees. *Journal of Ecology*, **92**, 758-766.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.

Higgins, S.I., Clark, J.S., Nathan, R., Hovestadt, T., Schurr, F., Fragoso, J.M.V., Aguiar, M.R., Ribbens, E. & Lavorel, S. (2003) Forecasting plant migration rates: managing uncertainty for risk assessment. *Journal of Ecology*, **91**, 341–347.

Higgins, S.I., Richardson, D.M. & Cowling, R.M. (2000) Using a dynamic landscape model for planning the management of alien plant invasions. *Ecological Applications*, **10**, 1833-1848.

Houghton, J.T., Ding, Y. & Griggs, D.J. (2001) Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change.). Cambridge University Press, Cambridge.

Howard, C.L., Mortimer, A.M., Gould, P., Putwain, P.D., Cousens, R. & Cussans, G.W. (1991) The dispersal of weeds: seed movement in arable agriculture. *Proceedings of the Brighton Crop Protection Conference, Weeds*, **2**, 821-828.

IPCC (2007) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment. Report of the Intergovernmental Panel on Climate Change. (ed. by S. Solomon, D. Qin, M. Manninget al). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Iverson, L.R., Prasad, A. & Schwartz, M.W. (1999) Modeling potential future individual tree-species distributions in the eastern United States under a climate change scenario: a case study with Pinus virginiana. *Ecological Modelling*, **115**, 77-93.

Iverson, L.R., Schwartz, M.W. & Prasad, A. (2004) How fast and far might tree species migrate in the eastern United States due to climate change? *Global Ecology & Biogeography*, **13**, 209-219.

Lischke, H., Ammann, B., Roberts, D.W. & Zimmermann, N.E. (2006) TreeMig: A forest-landscape model for simulating spatio-temporal patterns from stand to landscape scale. *Ecological Modelling*, **199**, 409-420.

Malcolm, J.R., Markham, A., Neilson, R.P. & Garaci, M. (2002) Estimated migration rates under scenarios of global climate change. *Journal of Biogeography*, **29**, 835-849.

McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models. 2nd edition,* London, Chapman and Hall.

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

Midgley, G.F., Hughes, G.O., Thuiller, W. & Rebelo, A.G. (2006) Migration rate limitations on climate change induced range shifts in Cape Proteaceae. *Diversity and Distributions*, **12**, 555-562.

Midgley, G.F., Thuiller, W. & Higgins, S.I. (2007) Plant species migration as a key uncertainty in predicting future impacts of climate change on ecosystems: progress and challenges. *Terrestrial ecosystems in a changing world*. (ed. by J. G. Canadell, D. E. Pataki and L. F. Pitelka), pp. 129-137. Springer-Verlag, New-York.

Müller-Schneider, P. (1986) Verbreitungsbiologie der Blütenpflanzen Graubündens. *Veröffentlichtes Geobotanisches Institut ETH Stiftung Rübel Zürich* **85**, 1-263.

Nathan, R. (2005) Long-distance dispersal research: building a network of yellow brick roads. *Diversity and Distributions*, **11**, 125–130.

Nathan, R. (2006) Long-Distance Dispersal of Plants. *Science*, **313**, 786-788.

Nathan, R. & Muller-Landau, H.C. (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology and Evolution*, **15**, 278-285.

Nathan, R., Perry, G., Cronin, J.T., Strand, A.E. & Cain, M.L. (2003) Methods for estimating long-distance dispersal. *Oikos*, **103**, 261–273.

Nathan, R., Sapir, N., Trakhtenbrot, A., Katul, G.G., Bohrer, G., Otte, M., Avissar, R., Soons, M.B., Horn, H.S., Wikelski, M. & Levin, S.A. (2005) Long-distance biological transport processes through the air: can nature's complexity be unfolded in silico? *Diversity and Distributions*, **11**, 131–137.

Nathan, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A. & Tsoar, A. (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology and Evolution*, **23**, 638-647.

Ostendorf, B., Hilbert, D.W. & Hopkins, M.S. (2001) The effect of climate change on tropical rainforest vegetation pattern. *Ecological Modelling*, **145**, 211-224.

Pearson, R.G. & Dawson, T.P. (2005) Long-distance plant dispersal and habitat fragmentation: identifying conservation targetsfor spatial landscape planning under climate change. *Biological Conservation*, **123**, 389-401.

Pitelka, L.F., Gardner, R.H., Ash, J., Berry, S., Gitay, H., Noble, I.R., Saunders, A., Bradshaw, R.H.W., Brubaker, L., Clark, J.S., Davis, M.B., Sugita, S., Dyer, J.M., Hengeveld, R., Hope, G., Huntley, B., King, G.A., Lavorel, S., Mack, R.N., Malanson, G.P., McGlone, M., Prentice, I.C. & Rejmanek, M. (1997) Plant migration and climate change. *American Scientist*, **85**, 464-473.

Randin, C.F., Dirnbock, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689-1703.

Ronce, O. (2001) Understanding plant dispersal and migration. *Trends in Ecology and Evolution*, **16**, 663.

Sarkar, P. (2000) A Brief History of Cellular Automata. *ACM Computing Surveys*, **32**, 80-107.

Scheller, R.M., Domingo, J.B., Sturtevant, B.R., Williams, J.S., Rudy, A., Gustafson, E.J. & Mladenoff, D.J. (2007) Design, development, and application of LANDIS-II, a spatial landscape simulation model with flexible temporal and spatial resolution. *Ecological modelling*, **201**, 409-419.

Silvertown, J., Holtier, S., Johnson, J. & Dale, P. (1992) Cellular Automaton Models of Interspecific Competition for Space - the Effect of Pattern on Process. *Journal of Ecology*, **80**, 527-534.

Soons, M.B., Heil, G.W., Nathan, R. & Katul, G.G. (2004) Determinants of long-distance seed dispersal by wind in grasslands. *Ecology*, **85**, 3056–3068.

Takahashi, K. & Kamitani, T. (2004) Effect of dispersal capacity on forest plant migration at a landscape scale. *Journal of Ecology*, **92**, 778–785.

Takenberg, O. (2003) Modeling long-distance dispersal of plant diaspores by wind. *Ecological Monographs*, **73**, 173–189.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., Ferreira de Siqueira, M., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427**, 145-147.

Thuiller, W. (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, **10**, 2020-2027.

Thuiller, W., Albert, C., Araujo, M.B., Berry, P.M., Guisan, A., Hickler, T., Midgley, G.F., Paterson, J., Schurr, F.M., Sykes, M.T. & Zimmermann, N.E. (2008) Predicting global change impacts on plant species distributions: future challenges. *Perspective in Plant Ecology, Evolution and Systematics*, **9**, 137-152.

Thuiller, W., Lavorel, S., Araujo, M.B., Sykes, M.T. & Prentice, I.C. (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8245-8250.

Vittoz, P. & Engler, R. (2007) Seed dispersal distances: a typology based on dispersal modes and plant traits. *Botanica Helvetica*, **117**, 109-124.

Ward, B.C., Scheller, R.M. & Mladenoff, D.J. (2004) Technical Report: LANDIS-II double exponential seed dispersal algorithm. University of Wisconsin-Madison, WI. http://www.landis-ii.org/documentation.

Willson, M.F. (1993) Dispersal mode, seed shadows, and colonization patterns. *Vegetatio*, **107-108**, 261-280.

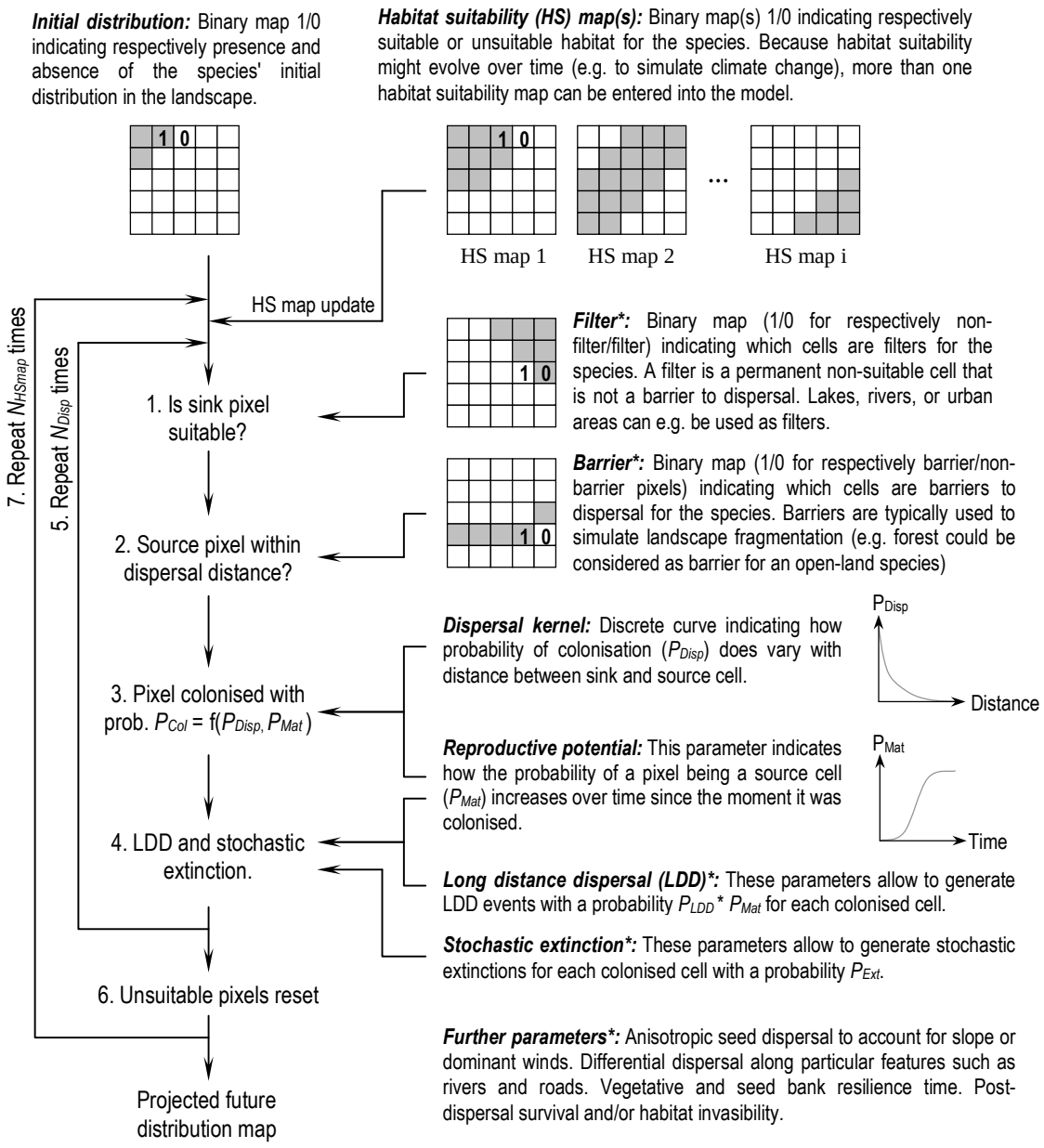
Woodward, F.I. & Beerling, D.J. (1997) The dynamics of vegetation change: health warnings for equilibrium 'dodo' models. *Global Ecology and Biogeography Letters*, **6**, 413-418.

Yemshanov, D. & Perera, A.H. (2002) A spatially explicit stochastic model to simulate boreal forest transitions: general structure and proprieties. *Ecological modeling*, **150**, 189-209.

Supplementary Material / Supporting Information

Appendix S1

Schematic view of the MIGCLIM model showing its most important parameters and where they act. Stars (*) indicate optional parameters. All parameters are modifiable to best fit a species' dispersal behaviour. Numbers 1-7 refer to those found in the main text. source cell = colonised pixel with reproductive ability, sink cell = potentially suitable but unoccupied pixel, LDD = long distance dispersal.



Appendix S2 Detail of P_{Disp} and P_{Mat} values used in the different simulations.

Values of increase in cell reproductive potential over time (P_{Mat}) used in Herb-type and Tree-type simulations. Note that in the Tree-type increase of reproductive potential, the cells have no reproductive potential before they reach 10 years. Once P_{Mat} = 1 it remains so unless the pixel becomes decolonised.

Herb-type simulations	
Time since colonisation [years]	value of P_{Mat}
1	0.01
2	0.08
3	0.50
4	0.92
5	1.00

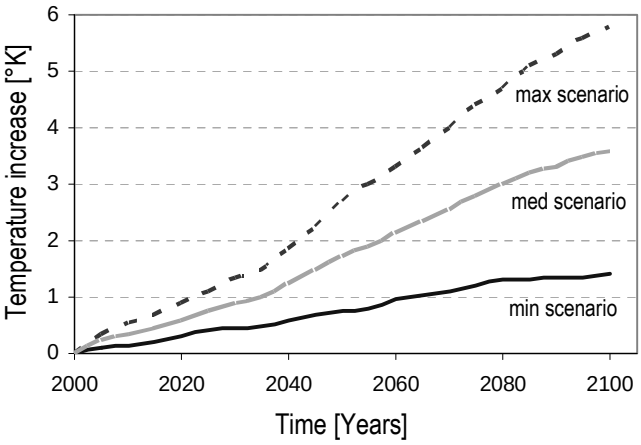
Tree-type simulations	
Time since colonisation [years]	value of P_{Mat}
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0.010
11	0.010
12	0.015
13	0.020
14	0.025
15	0.035
16	0.050
17	0.065
18	0.090
19	0.120
20	0.160
21	0.210
22	0.270
23	0.345
24	0.425
25	0.505
26	0.590
27	0.665
28	0.735
29	0.795
30	0.845
31	0.885
32	0.915
33	0.935
34	0.955
35	0.965
36	0.975
37	0.985
38	0.990
39	0.995
40	1.000

Values of P_{Disp} (multiplied by 100) used in each simulation. Simulations run at pixel sizes of 25 meters had their value of P_{Disp} for immediately adjacent pixels set to 0.99. This calibration was used in the absence of knowledge of the true value of "*Successful Seeds*" (Eq. 5) per surface unit. Simulations run at pixel sizes of 5 and 12.5 meters had their value of P_{Disp} adapted so that the density of "*Successful Seeds*" per surface unit remained the same as for the 25 meter pixel simulations. This ensured constant spread rate for the species regardless of pixel size. H = Herb-type, T = Tree-type.

simulation name	5m H	25m H	100m H	500m H	1000m H	100m T	500m T	1000m T
Pixel size	5 m		12.5 m	25 m		12.5 m	25 m	
Dist. from source cell [in pixels]	**Value of $P_{Disp} \times 100$**							
1	16	16	66	99	99	66	99	99
2	3.9	3.9	30	88	90	30	88	90
3	1.1	1.1	13	69	77	13	69	77
4	0.35	0.35	6.1	52	64	6.1	52	64
5	0.11	0.11	2.8	38	53	2.8	38	53
6	-	-	1.4	27	43	1.4	27	43
7	-	-	0.66	20	35	0.66	20	35
8	-	-	0.33	14	29	0.33	14	29
9	-	-	-	10	24	-	10	24
10	-	-	-	7.6	20	-	7.6	20
11	-	-	-	5.5	16	-	5.5	16
12	-	-	-	4.1	14	-	4.1	14
13	-	-	-	3.0	11	-	3.0	11
14	-	-	-	2.2	9.5	-	2.2	9.5
15	-	-	-	1.7	8.0	-	1.7	8.0
16	-	-	-	1.2	6.8	-	1.2	6.8
17	-	-	-	0.94	5.7	-	0.94	5.7
18	-	-	-	0.70	4.8	-	0.70	4.8
19	-	-	-	0.53	4.1	-	0.53	4.1
20	-	-	-	0.40	3.5	-	0.40	3.5
21	-	-	-	-	3.0	-	-	3.0
22	-	-	-	-	2.5	-	-	2.5
23	-	-	-	-	2.2	-	-	2.2
24	-	-	-	-	1.9	-	-	1.9
25	-	-	-	-	1.6	-	-	1.6
26	-	-	-	-	1.4	-	-	1.4
27	-	-	-	-	1.2	-	-	1.2
28	-	-	-	-	1.0	-	-	1.0
29	-	-	-	-	0.87	-	-	0.87
30	-	-	-	-	0.75	-	-	0.75
31	-	-	-	-	0.65	-	-	0.65
32	-	-	-	-	0.56	-	-	0.56
33	-	-	-	-	0.49	-	-	0.49
34	-	-	-	-	0.42	-	-	0.42
35	-	-	-	-	0.36	-	-	0.36
36	-	-	-	-	0.32	-	-	0.32
37	-	-	-	-	0.27	-	-	0.27
38	-	-	-	-	0.24	-	-	0.24
39	-	-	-	-	0.21	-	-	0.21
40	-	-	-	-	0.18	-	-	0.18

Appendix S3

The three temperature increase scenarios implemented in this study, based on the IPCC projections for the next 100 years (Houghton et al. 2001). The "max" (+ 5.8 °K) and "min" (+ 1.4 °K) scenarios correspond respectively to the maximum and minimum of the IPCC estimates while the "med" scenario (+ 3.6 °K) was obtained by averaging the two extreme scenarios. Temperature increase is not linear over time.



Appendix S4

Initial set of environmental variables used as input for the GLM stepwise selection procedure.

Variable	Units	Details	Reference
Mean annual daily solar radiation	kJ m^{-2} day^{-1}	Sum of daily solar radiations averaged over the period 1961-1990.	Zimmermann and Kienast (1999)
Mean annual temperature	°C	Average for period 1961-1990.	Zimmermann and Kienast (1999)
Temperature degree days	°C day $year^{-1}$	Sum of days with temperature > 0°C.	Zimmermann and Kienast (1999)
Snow cover	Days	Number of days with snow cover per year averaged for period 1980-2000.	Gurtz et al. (1999)
Slope	Degrees	Slope inclination	Anon (2004)

References

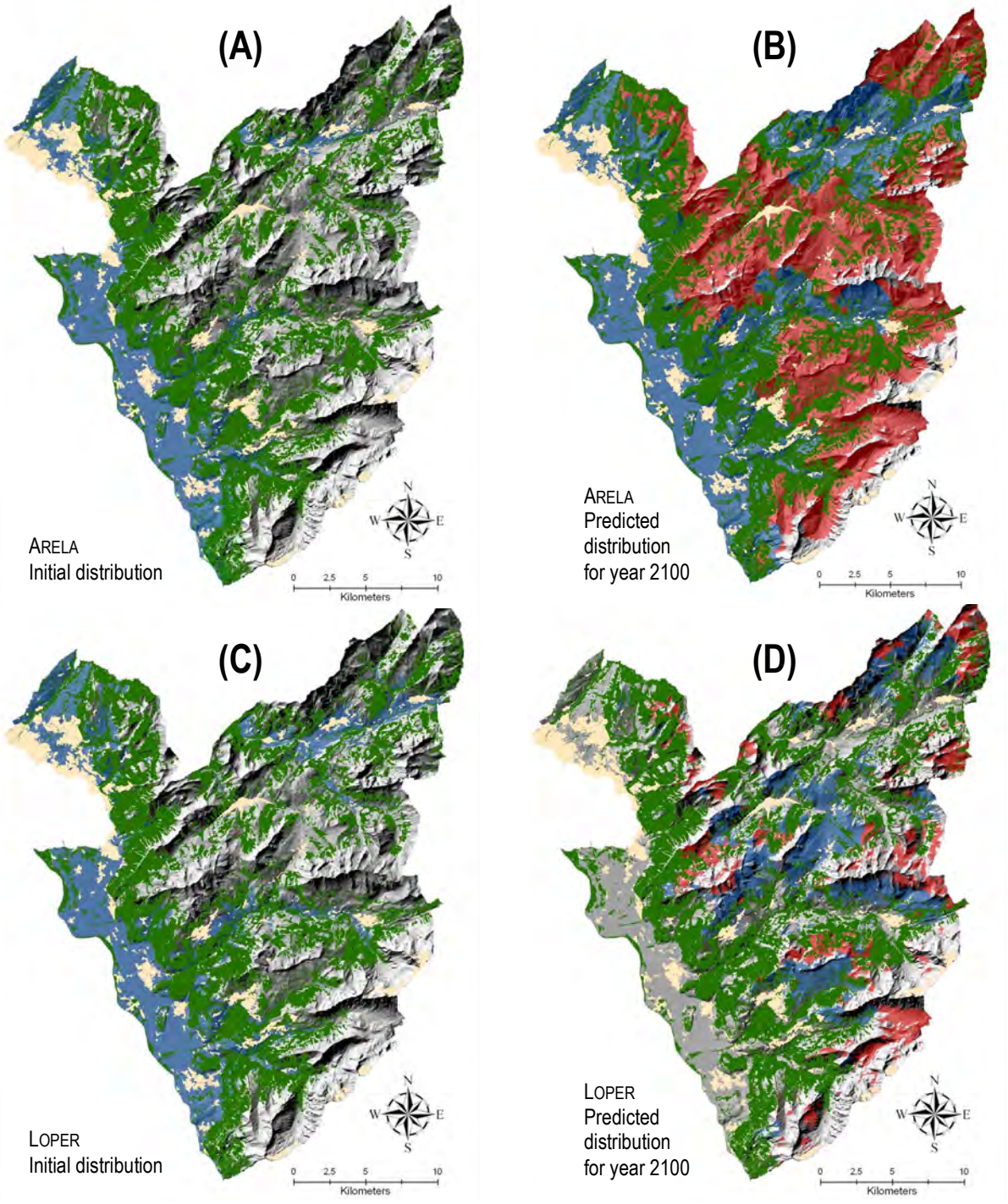
Anon (2004) ARCInfo version 9.0. Environmental Systems Research Institute, Redlands, CA, USA.

Gurtz, J., Baltensweiler, A. & Lang, H. (1999) Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins. *Hydrological Processes*, **13**, 2751–2768.

Zimmermann, N.E. & Kienast, F. (1999) Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. *Journal of Vegetation Science*, **10,** 469-482.

Appendix S5

Initial distribution of ARELA (A) and LOPER (C). Examples of predictions obtained for year 2100 with maximum temperature warming and 100m Herb dispersal scenario for ARELA (B) and LOPER (D) showing the potentially *colonisable* habitat (blue) and potentially *suitable* habitat (blue + red).



Initial distribution of ARELA (A) and LOPER (C); Predicted potentially colonisable habitat of ARELA (B) and LOPER (D) in 2100 under max temperature warming and 100m Herb dispersal scenario.

Potentially suitable but uncolonised surface for ARELA (B) and LOPER (D) in 2100 under max temperature warming and 100m Herb dispersal scenario.

Barriers to dispersal (forests).

Permanent unsuitable areas (filters).

Appendix S6

Explanation and mathematical justification of the optional pixel-size correction factor to ensure quasi-equal spread rate for simulations using different cell sizes with a same kernel of P_{Disp}.

Ideally, the MIGCLIM model is calibrated while knowing the production of "Successful Seeds" per surface unit, which allows to explicitly compute the value of P_{Disp} (probability of a pixel to be colonised given its distance to a source pixel) for any $pixel_j$ using the equation below (Eq. 5 in the present manuscript):

$$P_{Disp}(pixel_j) = 1 - (1 - P_{seed}(pixel_j))^{Successful\ Seeds} \qquad \text{(Eq. 5)}$$

What is important to note here is that the value of *Successful Seeds* is dependent upon the chosen pixel size because, for a same species, larger pixels should produce more successful seeds than smaller ones. In other words, the value of *Successful Seeds* used in Eq. 5 should always be proportional to pixel surface.

In practice, this means that if a P_{Disp} kernel is calibrated for a certain pixel size, then it cannot be used as such with another pixel size. If it is, then the number of *Successful Seeds* per surface unit is modified and so is the species dispersal rate. Thus, P_{Disp} kernels should be adapted to pixel size so that the value of *Successful Seeds* is proportional to the surface of a pixel (in biological term this means that the seed production per surface unit remains constant). This is the approach that was taken in the present study and ensures that a species dispersal rate is independent of the chosen pixel size.

Another option to make the spread rate of a species independent of pixel size is using an optional MIGCLIM parameter that allows adjusting the value of P_{Disp} for different pixel sizes than the one they were calibrated on. This pixel-size correction factor simply multiplies P_{Disp}, so that P_{Disp} then becomes "$P_{Disp} \times correction\ factor$". To work efficiently the value that must be used for this pixel-size correction factor is "1 / ratio of cell surfaces", as is demonstrated here-below (to simplify the notation, $P_{Seed}(pixel_i)$ is abbreviated as P_{Seed}):

$$P_{Disp} \times Correction\ Factor =$$

$$\left(1 - (1 - P_{Seed})^{Successful\ Seeds}\right) \times Correction\ Factor =$$

$$\left(1 - \left(1 - a_1 P_{Seed} + a_2 P_{Seed}{}^2 - a_3 P_{Seed}{}^3 + \Lambda\ \Lambda + a_{Successful\ Seeds} P_{Seed}{}^{Successful\ Seeds}\right)\right) \times Correction\ Factor =$$

Where a_1, a_2, … $a_{Successful\ Seeds}$ are the coefficients of the *Successful Seeds*th row of "Pascal's Triangle", a geometric arrangement of the binomial coefficients in which the triangle's n^{th} row gives the coefficients associated to each term of the developed form of $(x + y)^n$ or $(x - y)^n$.

The a_1 term of this triangle always corresponds to the number of the row of Pascal's Triangle, so that we can replace a_1 with *Successful Seeds* and re-write our expression as:

$$\left(1-\left(1-Successful\,Seeds\cdot P_{Seed}+a_2{P_{Seed}}^2-a_3{P_{Seed}}^3+\Lambda+a_{Successful\,Seeds}{P_{Seed}}^{Successful\,Seeds}\right)\right)\times Correction\,Factor=$$

This part of the expression quickly becomes negligible as P_{Seed} becomes << 1.

$$\left(Successful\,Seeds\cdot P_{Seed}-a_2{P_{Seed}}^2+a_3{P_{Seed}}^3-\Lambda-a_{Successful\,Seeds}{P_{Seed}}^{Successful\,Seeds}\right)\times Correction\,Factor\cong$$

Since P_{Seed} is a small number that quickly becomes << 1, the part of the expression highlighted above becomes negligible in comparison to *Successful Seeds* × P_{Seed} so that we can rewrite our expression as:

$$Successful\,Seeds\cdot P_{Seed}\cdot Correction\,Factor=$$

Because the number of "*Successful Seeds*" should be proportional to the surface of a pixel, we can re-write our expression above as:

$$Successful\,Seeds\,Per\,Surface\,Unit\cdot surface\cdot P_{Seed}\cdot Correction\;Factor$$

It becomes apparent that if the value of *Correction Factor* is inversely proportional to *surface*, then the *surface* term is canceled-out of the expression. If we now want to cancel out the difference between pixel surfaces rather than surface as a whole, then the *Correction Factor* must be set equal to "1 / ratio of pixel surfaces" (e.g. the surface ratio between 25 m and 12.5 m pixels is $25^2/12.5^2 = 4$, therefore the value of the *Correction Factor* must be set to 0.25).

Thus, setting the value of the *Correction Factor* to "1 / ratio of cell surfaces" ensures a quasi-equal spread rate among simulation run at different pixel sizes with a same kernel of P_{Disp} values.

The table below shows an example of a 10 pixel kernel (values of P_{Disp}) developed for three different pixel sizes: 25 m (column A), 12.5 m (column B) and 5 m (column C). The values of P_{Disp} were obtained by using Eq. 5 and a fixed density of "Successful Seeds" production per surface unit. In other words, they were developed by accounting for the fact that a 25m pixel should produce 4 times more "Successful Seeds" than a 12.5m pixel ($25^2/12.5^2 = 4$) and 25 times more "Successful Seeds" than a 5m pixel ($25^2/5^2 = 25$).

Columns D and E respectively show the ratio of these kernels: Column D presents the ratio between columns B/A and column E the ratio of columns C/A. As can be observed, these ratios quickly converge towards 0.25 (which is the ration of $12.5^2/25^2 = ¼ = 0.25$) and 0.04 (which is the ration of $5^2/25^2 = 1/25 = 0.04$). This shows in practice that the use of "1 / cell surface ratio" as correction factor will produce almost the same results as considering a constant seed production per surface unit (and recalibrating the values of P_{Disp} for each pixel size), and hence result in a very similar spread rate of the species independently of the chosen cell size.

Example of 10 pixel kernel					
	A	B	C	D	E
Distance from source cell [in pixels]	25m pixel	12.5m pixel	5m pixel	Ratio 12.5/25	Ratio 5/25
1	0.9900	0.6600	0.1600	0.67	0.16
2	0.8000	0.3300	0.0620	0.41	0.08
3	0.5100	0.1600	0.0280	0.32	0.06
4	0.3000	0.0840	0.0140	0.28	0.05
5	0.1700	0.0440	0.0072	0.27	0.04
6	0.0920	0.0240	0.0039	0.26	0.04
7	0.0510	0.0130	0.0021	0.25	0.04
8	0.0290	0.0073	0.0012	0.25	0.04
9	0.0160	0.0041	0.0006	0.25	0.04
10	0.0094	0.0024	0.0004	0.25	0.04

Finally we would like to emphasize that this correction factor is optional and should be used only if one cannot or does not want to compute the seed production per surface unit. If the seed production per surface unit can be computed, then we recommend using this information in Eq. 5 to re-compute new values of P_{Disp} (as done in the present manuscript) as this is more meaningful from a biological point of view and will produce slightly more accurate results.

2.2

Seed dispersal distances: a typology based on dispersal modes and plant traits.

Pascal Vittoz[1] and Robin Engler[2]

This manuscript has been published in 2007 *Botanica Helvetica*, **117**, 109-124.

DOI: 10.1007/s00035-007-0797-8

[1] Faculty of Geosciences and Environment, University of Lausanne, CH-1015 Lausanne, Switzerland.

[2] Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland.

CONTRIBUTION TO THE PROJECT:
I collaborated with PV to the literature review. I created the figure and participated in the reviewing process of the manuscript.

Abstract

The ability of plants to disperse seeds may be critical for their survival under the current constraints of landscape fragmentation and climate change. Seed dispersal distance would therefore be an important variable to include in species distribution models. Unfortunately, data on dispersal distances are scarce, and seed dispersion models only exist for some species with particular dispersal modes. To overcome this lack of knowledge, we propose a simple approach to estimate seed dispersal distances for a whole regional flora. We reviewed literature about seed dispersal in temperate regions and compiled data for dispersal distances together with information about the dispersal mode and plant traits. Based on this information, we identified seven "dispersal types" with similar dispersal distances. For each type, upper limits for the distance within which 50% and 99% of a species' seeds will disperse were estimated with the 80th percentile of the available values. These distances varied 5000-fold among the seven dispersal types, but generally less than 50-fold within the types. Thus, our dispersal types represented a large part of the variation in observed dispersal distances. The attribution of a dispersal type to a particular species only requires information that is already available in databases for most Central European species, i.e. dispersal vector (e.g. wind, animals), the precise mode of dispersal (e.g. dyszoochory, epizoochory), and species traits influencing the efficiency of dispersal (e.g. plant height, typical habitats). This typology could be extended to other regions and will make it possible to include seed dispersal in species distribution models.

Keywords: Anemochory, anthropochory, autochory, hydrochory, plant migration, zoochory.

Introduction

Plant dispersal has attracted scientists since long ago (Darwin 1859; Schmidt 1918; Ridley 1930; Müller-Schneider 1983) and is particularly relevant with relation to human-driven environmental changes. For example, the survival of plant metapopulations in fragmented landscapes strongly depends on their dispersal potential (Fischer et al. 1996; Couvreur et al. 2004; Soons and Ozinga 2005), and the predicted global warming will require considerable migration rates for plant species to remain under similar climatic conditions (Malcolm et al. 2002). Nevertheless, most models attempting to predict future plant distributions did not include dispersal, considering it as unlimited (Guisan and Theurillat 2000; Thuiller et al. 2005). Even without constraints on seed dispersal, these models already predict local extinctions, e.g. for isolated populations in mountains (Guisan and Theurillat 2000; Dirnbock et al. 2003; Thuiller et al. 2005). The actual extinction rates might be even higher if plant species cannot keep pace with rapid climate change due to limited dispersal. A more precise assessment of plant species extinction risk thus calls for the incorporation of plant dispersal potential (Pitelka et al. 1997; Davis et al. 1998; Ronce 2001).

Many studies have measured or estimated dispersal distances of plants in the field (Schneider 1935; Stöcklin and Bäumler 1996; Jongejans and Telenius 2001), and several mathematical models have been developed to estimate these distances (Tackenberg et al. 2003; Mouissie et al. 2005a; Nathan et al. 2005; Soons and Ozinga 2005). However, all of these studies have considered only a limited number of species or dispersal vectors. No dispersal distance data exist for a complete regional flora. Müller-Schneider (1986) reviewed dispersal vectors for the entire flora of Graubünden (East of Switzerland), but his work includes only few dispersal distances, most of which stem from anecdotal observations. Likewise, Bonn and Poschlod (1998) and Bonn (2004) wrote important syntheses on seed dispersal in Central Europe, but dispersal distances were only provided for a few dispersal vectors, mainly from anecdotal observations. It is thus currently impossible to conduct an assessment of the extinction risk of plant species under landscape fragmentation or global warming that would take dispersal into account.

The distance over which plants disperse seeds depends on plant traits as well as environmental conditions and varies strongly in time and space. This variability can be represented by a dispersal curve (dispersal kernel), which gives the proportion of seeds reaching a given distance (Mouissie et al. 2005a). However, it would be highly time consuming, if not impossible, to determine dispersal kernels for each species of a region. Thus, a simplified approach is needed to estimate dispersal distances for a whole regional flora. For example, if dispersal curves could be classified into a limited number of types with similar dispersal distances, and if plant species could be attributed to these "dispersal types" based on generally available plant traits, it would be possible to estimate dispersal kernels for all of them.

In this paper, we develop such an approach for the Swiss flora based on an extensive review of seed dispersal literature. We propose a typology of dispersal curves that can be applied to most Swiss and Central European plants. This typology could be extended to other regions and could be used to account for dispersal distances in species distribution models, enabling refined extinction risk assessments to be made for large numbers of species.

Methods

Plant dispersal is generally achieved through seeds. These can be enclosed in fruits or larger structures (usually called "diaspores"), but for the sake of simplification, the term "seed" will be used here as a general denomination.

Data for seed dispersal distances were compiled by reviewing a large proportion of available literature from Switzerland and other European countries, including monographies (Müller-Schneider 1983, 1986), reviews and research articles. Swiss species or close relatives were considered first priority, since our aim was to develop a typology for this region. However, other species were included when data available for Swiss species were insufficient to assess dispersal distances for certain dispersal modes (see below). The complete data set (ca. 300 values) is presented in Appendix 1. Species nomenclature follows Aeschimann et al. (1996) for the Swiss species.

The data set proved to be very heterogeneous. A small proportion of the distances had been determined through experiments, detailed field observations of seed or seedling distributions, or mathematical models. In such cases, it is often possible to calculate a dispersal kernel. However, most of the available data represent isolated and often anecdotal observations, from which a precise dispersal kernel cannot be derived. Some of these isolated observations clearly represented long-distance dispersal events (LDD), i.e. extreme values reached only by a very small minority of seeds. We therefore classified the data into three categories: (1) mean, mode or median values, (2) maximum values (99th percentiles of distribution kernels) and (3) values for LDD (clearly above the potential dispersal of 99% of the seeds). LDD values were excluded from the further analysis of the data.

Our typology of dispersal curves was based on the dispersal modes recognised by Müller-Schneider (1983). The English translation of Müller-Schneider's German terminology generally follows Bonn et al. (2000). Müller-Schneider's (1983) classification of dispersal modes is primarily based on the dispersal vector (wind, water, animals, etc.), with additional subdivisions for the differing ways in which seeds are released and transported (e.g. on the fur or after ingestion by animals). Additional subdivisions were made for dispersal modes whose efficiency clearly depends on supplementary factors: plant height, pappus efficiency and environing vegetation structure for anemochory, and vector size for zoochory. Of the numerous possible subdivisions, only those considered most relevant were retained for our classification, as explained in the next section. This yielded a total of 21 refined dispersal modes (Tab. 1).

Each dispersal distance in our data set was attributed to a dispersal mode, which was either the mode for which the distance had been determined (if mentioned in the original study) or the main dispersal mode of the species according to Müller-Schneider (1986). For species with more than one dispersal mode, distances that could not be clearly related to one of the modes were excluded from further data analysis. Dispersal types were then defined by grouping together dispersal modes with similar dispersal distances. This was done graphically by plotting the mean and maximal distances for each dispersal mode and identifying modes for which distances were in the same order of magnitude (Fig. 1).

Tab. 1. Dispersal distances for seven dispersal types, estimated as the upper limits of the distances within which 50% and 99% of the seeds of a plant population are dispersed. Note that actual dispersal distances will usually be lower than those given here (cf. Fig. 1). The dispersal distances were estimated from the 80th percentile of the data compiled in Fig. 1 as well as additional qualitative information as explained in the text ('Dispersal modes and evaluation of published dispersal distances'). The dispersal modes included in each dispersal type are indicated; they are based on dispersal vectors (categories in parentheses) and plants traits that influence the efficiency of dispersal.

Type	Dispersal distances [m]		Corresponding dispersal modes
	50%	99%	
1	0.1	1	Blastochory (autochory) Boleochory (anemochory) for species < 30 cm Ombrochory (hydrochory)
2	1	5	Ballochory (autochory) Cystometeorochory (anemochory) Chamaechory (anemochory) for fruits in grassland Boleochory (anemochory) for species > 30 cm
3	2	15	Pterometeorochory (anemochory) for herbs Myrmecochory (zoochory) Cystometeorochory (anemochory) ferns, Orchidaceae, Pyrolaceae, Orobanchaceae in forest Trichometeorochory (anemochory) in forest or little efficient plumes Epizoochory (zoochory) for small mammals
4	40	150	Chamaechory (anemochory) for seeds on snow or dry inflorescence Pterometeorochory (anemochory) for trees Dyszoochory (zoochory) for seeds not stocked and dispersed by small animals
5	10	500	Trichometeorochory (anemochory) in openland with efficient plumes Cystometeorochory (anemochory) ferns, Orchidaceae, Pyrolaceae, Orobanchaceae in openland
6	400	1500	Dyszoochory (zoochory) for seeds stocked by large animals Endozoochory (zoochory) for seeds eaten by birds and large vertebrates Epizoochory (zoochory) by large mammals
7	500	5000	Agochory (anthropochory)

Finally, we estimated upper limits of the distances, within which 50% and 99% of the seeds would disperse, by using the 80th percentiles of the available mean, mode or median values and of the maximum values. Results were rounded to one significant digit to reflect their approximate nature. Our aim was to provide a conservative estimate of the dispersal constraint experienced by most species belonging to a dispersal type. Therefore we did not take the average of the published values (Fig. 1), but rather the 80th percentile of the distribution, as this allowed us to exclude the most extreme values. In some cases, a comparison of the results with qualitative information from the literature or with the authors' experience indicated that the available data were not quite representative for a certain dispersal type; values were then adjusted to obtain more realistic estimates. Such decisions are explained in the next section of the text for the individual dispersal modes.

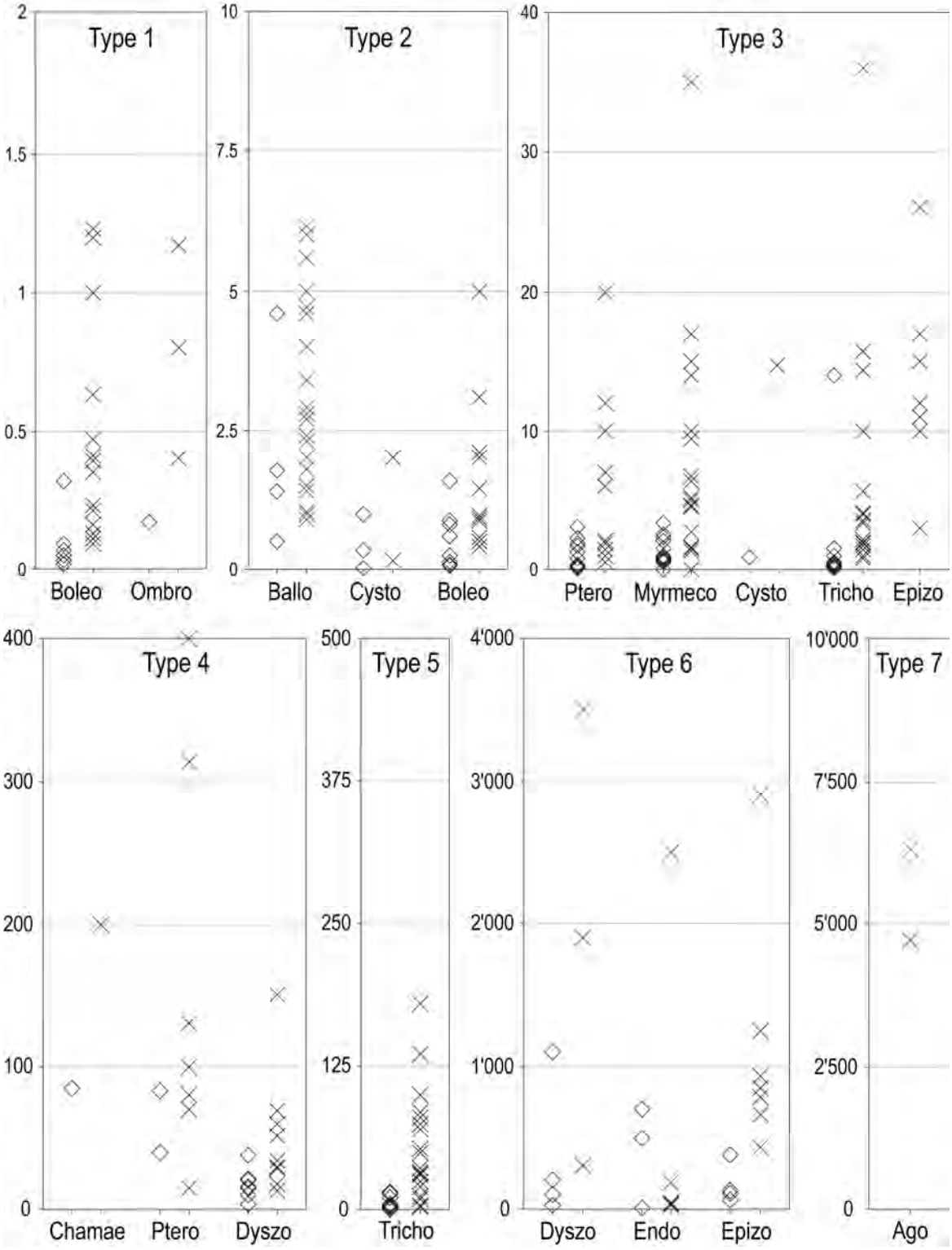


Fig. 1. Distribution of the dispersal distances found in literature (Appendix 1) for each dispersal mode, and subdivision of the data set into seven dispersal types. Diamonds are for mean, median or mode values, and crosses for 99% or maximum values (without long-distance dispersal). Four retained maximum values of type 5 are outside of the graph: 1714 m, 2112 m, 2194 m and 3673 m. See Table 1 for definitions of the dispersal modes.

Dispersal modes and evaluation of published dispersal distances

Autochory

Autochorous plants disperse seeds without the help of an external vector. As a result, dispersal is limited to very short distances.

In **blastochory**, the stem of the plant grows or crawls on the ground to deposit the seeds as far as possible from the mother plant (e.g. *Cymbalaria muralis*, *Polygonum aviculare*, *Veronica hederifolia*; Müller-Schneider 1983). No data were found in the literature, but since the dispersal distance corresponds to the length of the stem, although species-specific, it is mostly very short and blastochory can hence be classified as type 1 (Tab. 1). This dispersal mode is, however, frequently completed by another one (Müller-Schneider 1986).

In **ballochory,** the explosion of the fruit ejects the seeds (ballistichory, ballistic dispersal**)**. This explosion may be due to the turgescence of tissues (*Impatiens sp., Cardamine sp.*) or the tension between cells or different cell layers when the fruit is drying (*Viola sp., Vicia sp., Lotus sp.*). Published values are scattered and very variable (maximum 0.89-6.2 m; Fig. 1). Ballochory is classified in dispersal type 2 (Tab. 1).

Two further dispersal modes are **barochory** (seeds fall from the plant) and **herpochory** (seeds creep on the soil by the movement of organs in a succession of dry and wet conditions). However, since these strategies are not very efficient and always combined with other dispersal modes (Müller-Schneider 1986), they were not retained here.

Anemochory

Anemochorous seeds are dispersed by wind, often with the help of specific organs. This dispersal vector is the most studied as it is easily observable and measurable, at least over short distances (e.g. Bullock and Clarke 2000; Jongejans and Telenius 2001). Moreover, it relies on physical processes that can be translated into models (Tackenberg et al. 2003; Nathan et al. 2005; Soons and Ozinga 2005). Anemochory is subdivided according to the organs used to slow down the falling of seeds.

An air filled structure lightens small seeds in **cystometeorochory** (balloon-like). This dispersal mode is little studied. Maximum calculated distances are below 2 m (Soons and Ozinga 2005), but extreme values were measured up to 80 m for *Calluna vulgaris* (Bullock and Clarke 2000). This mode is certainly less efficient in forests, as wind is weaker, but it seemed useless to subdivide these already low values and thus cystometeorochory as a whole was attributed to type 2.

The tiny seeds of Orchidaceae, Pyrolaceae and Orobanchaceae also have a low falling velocity (0.2-0.31 m/s for Orchidaceae; Müller-Schneider 1986). But only a calculated dispersion distance is available (median 0.95 m and 99-percentile 14.7 m for *Cephalanthera damasonium*, Soons and Ozinga 2005). However, because it is thought that very light seeds (<0.05 mg), even without corresponding adaptation for anemochory, are as efficient in wind as plumed seeds (Bonn and Poschlod 1998; Greene and Calogeropoulos 2002), we decided to classify these plant families with trichometeorochory in type 5 in openland but decreased to type 3 for forest species (Tab. 1). Fern spores can be included in cystometeorochory as well, but no data exist on their dispersal capacity except a calculated distance of 330 km for *Lycopodium sp.* based on its very low falling velocity (1.8 cm/s; Schmidt 1918). This value seems exaggerated and in the absence of a more precise value, we attributed the ferns to the same types as orchids.

Plumed seeds are more efficient for wind dispersal. In **trichometeorochory**, seeds are completed with a hairy structure (e.g. pappus) to reduce falling velocity. These organs have very variable efficiency, however, with falling velocity varying from 8 cm/s for *Epilobium angustifolium* to 165 cm/s for *Pulsatilla alpina* (Müller-Schneider 1986). With an arbitrary separation at 30 cm/s, on the basis of our own observations, we distinguished species with less efficient plumes from those with efficient plumes (long plumes for small seeds). The first group has maximum distances between 1-15.7 (36) m, corresponding to type 3, and the second mainly between 20 and 179 m (Fig. 1). However, because some species have much higher calculated potentiality (e.g. up to 3600 m for *Typha latifolia*; Soons and Ozinga 2005), we retained intermediate values and assigned trichometeorochory to type 5. Forest species were classified with trichometeorochory for less efficient plumes (type 3).

In **pterometeorochory** (or pterochory), seed dispersal is improved through wings. Trees are frequent in this category, but herbs are present as well, with a generally higher falling velocity. Because tree seeds are often large and easy to find, many available maximum dispersal distances are to be classified as LDD (e.g. Müller-Schneider 1983). Reviewed maximum distances ranged mainly between 80-314 m for trees and 1-12 m for herbs (Fig. 1). Pterometeorochory was thus classified as dispersal type 3 for herbs and type 4 for trees (Tab. 1).

A much less studied dispersal mode is **chamaechory**, with diaspores rolling on the ground pushed by the wind. This diaspore can be either a circular-shaped fruit (*Colutea arborescens, Astragalus alpinus*), the fruit with calyx (*Anthyllis vulneraria*) or the complete, dry, inflorescence (synaptospermie of *Eryngium campestre, Carlina acaulis*). Chamaechory is especially common and efficient in steppes where nothing hampers dispersal (Müller-Schneider 1983), but it also occurs in mountains, with small seeds on snow (e.g. *Saxifraga bryoides, S. exarata, Sempervivum montanum*). The only available data are from Greene and Johnson (1997), who observed *Betula alleghaniensis* seeds and calculated a possible dispersion of 38 m for spherical 1mg-seeds on snow. Dispersal is usually restricted because seeds get stuck in irregularities. For chamaechory, we retained dispersal distance type 2 for fruits in grassland and type 3 for seeds on snow or carried by dry inflorescences (Tab. 1), but supplementary data would be necessary to get more precise values.

Boleochory (semachory) is another mode used by anemochorous plants. The small seeds without particular features are spread when the fruit is shaken by wind. At maturity, the stem of such plants is often rigid but elastic and sways in the wind, acting like a catapult. As animals or others may shake the capsules as well, some classify this mode independently (semachory; Bonn et al. 2000). Although small, the seeds are dense and have a high falling velocity (1.2-5 m/s; Müller-Schneider 1983; Tackenberg 2001). Consequently, Soons and Ozinga (2005) calculated very short dispersal distances, generally <0.5 m, but without considering the catapult effect. Yet, this effect is certainly important, as measured distances sometimes exceed 10 m and are always higher than calculations for the same or close species. Since the catapult effect strongly depends on the stem size, we distinguished small species (<30 cm) whose seeds rarely go beyond 1 m (type 1) from taller species (>30 cm), whose seeds may reach up to 3-5 m (type 2).

Hydrochory

Water can disperse seeds in various ways. In wetland plants, seeds are often light enough to float and move on rivers, lakes or ponds (**nautochory** of *Alisma plantago-aquatica, Carex flava, C. elata, Iris pseudacorus, Sparganium sp.*). Some seeds can float and survive for one year or

more (Müller-Schneider 1983). Similarly, running water may carry many different types of seeds with heavy rains (**bythisochory**), sometimes to rivers and down to lowland areas. Bythisochory is complementary to other dispersal modes and randomly affects many different species dwelling on slopes. It is through this vector that high mountain species are frequently observed on gravel areas along rivers (Bill et al. 1999). Although the dispersal distances may be important, we did not attribute dispersal types to hydrochorous dispersal modes because distances are highly unpredictable and never documented. Moreover, nautochory is geographically limited and the bythisochory downslope restricted.

Rain may contribute to disperse seeds through the shock generated by the rain droplets hitting the fruits (**ombrochory**). Some species (e.g. *Caltha palustris, Veronica serpyllifolia, Prunella vulgaris*, *Thlaspi perfoliatum*) have developed fruit shapes and elastic fruitstalks in order to use this energy to eject seeds. Very few measurements are published for ombrochory, but they are all below or close to 1 m (type 1).

Zoochory

Animals are frequent and efficient vectors of dispersal, either voluntary when foraging or involuntary when carrying seeds on their fur or in their guts. Even though zoochory has often been observed and studied, estimating dispersal distances nevertheless remains difficult, as they highly depend on the disperser's behaviour. Zoochory can be split into four subcategories.

Many seeds are foraged as food by animals, which sometimes hide them as stock for the winter and forget about them, or lose them during transport (**dyszoochory** or dysochory). Vectors are mainly rodents or birds, and the dispersal distance is thus strongly dependent on the vector size. Small rodents, like voles or mice (*Clethrionomys sp., Microtus sp., Apodemus sp.*), generally disperse seeds less than 30 m (Cain et al. 1998; Xiao et al. 2004), and squirrels (*Sciurus sp.*) a little farther. In most cases, small birds disperse seeds by chance when feeding, for example when tits or woodpeckers are looking for a convenient place to break a nut. The rare available data do not exceed 60 m. However, some larger species are more efficient dispersers by hiding fruits for winter stocks. The most famous examples are the nutcracker (*Nucifraga carcyocatactes*; Müller-Schneider 1986; Mattes 1992) and the jays (*Garrulus glandarius*; Müller-Schneider 1949; Kollmann and Schill 1996). The literature contains different data, but those are unfortunately too often extreme values (Mattes 1992), and most of the seeds are probably hidden within a few hundred meters. We thus retained type 4 when the vector of dyszoochory is a small animal and type 6 when seeds are stocked by a large animal (Tab. 1).

A particular case of dyszoochory is **myrmecochory**, or dispersal by ants. Generally interested by the elaiosome, a fatty appendix of the seeds, ants transport the seeds before eating the elaisome but leaving the rest of seed untouched and still able to germinate. Seeds may be used as building material for their nest as well, without loosing their germination potential (Müller-Schneider 1963; Cherix 1981). This dispersal mode has been extensively studied in the world, but only rarely are distances available for European plants, and they rarely exceed 10 m. Some exceptional observations nevertheless give values up to 70 m (Müller-Schneider 1983; Bonn and Poschlod 1998), and myrmecochory was hence classified as type 3.

Animals are important dispersal vectors when eating fruit or even the complete plant (Janzen 1984), and seeds go undamaged through their gut (**endozoochory**). Many authors have studied the survival of the seeds through vertebrate guts and the importance of this

vector (see Janzen 1984; Pakeman 2001). As the consumer can be anything from a worm to a snail, mammal or bird, dispersal distance is very dependent on its size and mobility. No data exist for small animals and they are scarce and mostly anecdotal for larger ones such as birds or foxes. Models based on seed-retention time is a possibility for getting dispersal distance estimates, but they are still rare (e.g. Hickey et al. 1999; Vellend et al. 2003), and seed-retention time depends on seed and animal species (Bonn 2004; Mouissie et al. 2005b). Moreover, these models usually calculate linear distance, but animals generally live in a limited territory and do not move linearly. We chose type 6 to translate potential dispersal by large mammals or birds (Tab. 1).

Seeds are also frequently transported by animals in fur (**epizoochory).** This is partly the result of specific structures, with seeds or fruits bearing hooks or glandulous hairs (*Galium aparine, Arctium sp., Saxifraga tridactylites*,...) but seeds without an appendix can attach to fur as well (e.g. Fischer et al. 1996; Mouissie et al. 2005a; Römermann et al. 2005). Observations in natural conditions are rare and most of the data are from retention time measurements with sheep, cattle or dummies (e.g. Fischer et al. 1996; Mouissie et al. 2005a). Although small rodents may disperse seeds as well (Kiviniemi and Telenius 1998), the most efficient epizoochory is obtained with taller animals. The maximum distance calculated by models based on seed retention time in fur is between 435-1242 m, but can be longer with sheep whose long and curled wool is particularly efficient at retaining seeds (Fischer et al. 1996; Mouissie et al. 2005a). The habitual dispersal distance is thus estimated as type 6 for epizoochory with large animals, but much longer distances can occasionally be achieved by sheep during transhumance (Fischer et al. 1996).

Anthropochory

Seed dispersal by humans certainly always occurred, but it strongly increased during the last centuries, and became particularly important a few decades ago with the market globalisation and the intercontinental transport of goods (e.g. Hodkinson and Thompson 1997; Tinner and Schumacher 2004).

Müller-Schneider (1983, 1986) distinguished three modes of anthropochory: plants or seeds being sold for agriculture and gardening (**ethelochory**), seeds being involuntarily mixed with the previous ones (**speirochory**), or seeds travelling hidden in goods, cars, soil under soles, with hay, etc. (**agochory**). All three means can potentially lead to very long dispersal distances and are, for example, responsible for the advent of neophytes in Switzerland and Europe. But while ethelochory and speirochory mostly concern urban and cultivated areas, agochory is probably more important in natural or semi-natural ecosystems. Seed dispersal distance through anthropochory is strongly dependent on the type of human activity but, in general, agricultural activities are the most susceptible to spreading seeds in semi-natural ecosystems due to movements between fields or meadows (McCanny and Cavers 1988). We can thus limit most of the dispersal distance to the approximate size of a farming property (type 7).

Dispersal types and estimated dispersal distances

Despite the heterogeneous origin of the data compiled here, dispersal distances for individual dispersal modes proved to be rather consistent, mostly belonging to the same order of magnitude. Across the entire data set, maximal dispersal distances ranged between 0.09 and 6300 m (LDD excluded), corresponding to a factor of 70'000 between the highest and lowest

value. After classification into dispersal types, this variation was reduced to a factor of 10 for type 1, 40 for type 2, 70 for type 3, 20 for type 4, 1700 for type 5, 200 for type 6, and 1 for type 7 (very few data). This variability within types may still seem considerable, but it is small compared to the 5000-fold difference in dispersal distances between types 1 and 7. Furthermore, the high value for type 5 (trichometeorochory with efficient plumes) reflects the high variability of pappus efficiency in this category and the high variability found within species (e.g. *Taraxacum officinale*). The typology presented here thus expresses a large part of the variation in seed dispersal distances. Accordingly, attributing species to dispersal types makes it possible to describe interspecific variation in dispersal capacity.

The estimated distances in Table 1 do of course not represent the dispersal kernel of one single plant, nor even the mean pluri-annual dispersal kernel of a particular plant population. They were estimated as the upper limits (80th percentile) of the dispersal distance values (Fig. 1), meaning that they represent the dispersal potential of the plant species grouped into a dispersal type. Most plant populations will disperse over smaller distances than those indicated in Table 1, but data and models indicate that they could potentially disperse 50% or 99% of their seeds inside the retained distances. Estimating upper limits to dispersal, rather than average distances, is justified when dispersal is included as a possible constraint to species survival in predictive models of species distributions. In this case, upper dispersal limits yield a constraint that holds for all species of a certain dispersal type. This ensures that dispersal constraints will not be overestimated.

Alternative dispersal modes

Multiple dispersal vectors

About 40% of the species considered by Müller-Schneider (1986) have two or more dispersal modes. The species can either use them alternatively depending on the available vector (*Picea abies* is anemochorous or dyszoochorous with the red squirrel or some birds) or on its phenology (*Urtica dioica* is anemochorous and avoided by animals when green but grazed and endozoochorous once dry), or it can rely on them successively to improve dispersal (*Leucojum vernum* is firstly blastochorous and lately myrmecochorous; Müller-Schneider 1983).

If the most obvious dispersal mode can often be inferred from the seed or fruit morphology, finding out what the alternative dispersal modes of a species are generally requires precise observations. For example, *Campanula rotundifolia* and *Primula elatior* are considered endozoochorous by Müller-Schneider (1986) but not *Campanula scheuchzeri* or *Primula veris*, which are only described as boleochor species. This difference, probably incorrect, strongly affects their dispersal potential, as endozoochory is much more efficient than boleochory (Tab. 1), and shows the gaps in our attainments.

Recent results showed that this problem appears with other dispersal modes too. Tackenberg et al. (2003) modelled wind dispersion of seeds on the basis of their falling velocity and release height. They concluded that some species normally not considered as anemochorous could be as efficient as species traditionally thought-of as wind dispersed. Another example is given by Higgins et al. (2003), who demonstrated that a 7.8 g *Carya glabra* nut is able to disperse 647 m if uplifted by strong winds. Similarly, epizoochory concerns more species than what diaspore morphology indicates, and many plumed seeds for anemochory or smooth seeds are transported as well (Fischer et al. 1996; Couvreur et al. 2004; Mouissie et al. 2005a; Römermann et al. 2005).

When multiple vectors are recognized, it is logical to classify the species into the dispersal distance type corresponding to the most efficient one (e.g. dyszoochory for *Picea abies* or endzoochory for *Campanula rotundifolia*). But this can not consider the unsuspected supplementary vectors.

Long-distance dispersal (LDD) and Reid's paradox

The inadequacy between the dispersal potential of plants and their post-glacial recolonisation, also known as "Reid's paradox" (Clark et al. 1998), is an issue that has been recognized for a long time (Reid 1899; Skellam 1951; Cain et al. 1998). Lang (1994) calculated the migration rate of anemochorous trees through Europe and found per-generation travel distances of 0.5-5 km for *Tilia sp*, 1.2-9 km for *Abies alba*, 10-20 km for *Acer sp*. or 15-60 km for *Pinus sylvestris*. This is much higher than the 200 m considered in table 1 for 99th percentile. Similarly, dyszoochorous species with an estimated potential dispersal of 1 km (Tab. 1) showed post-glacial colonisation rate of 2.2-15 km per generation for *Quercus sp*. or 7-14 km for *Fagus sylvatica* (Lang 1994). However, as was recently found for *Fagus sylvatica* (Magri et al. 2006), it is possible that those recolonisation rates are overestimated because some glacial refugia remain yet unknown (Clark et al. 1998; Stewart and Lister 2001; Pearson 2006).

Recent data for invasive species show similar high rates of spread for many species. Pyšek and Hulme (2005) listed 16 species with colonisation superior to 1 km/y for long-distance dispersal, with a maximum of 167 km yr^{-1}. They showed that the rate of spread may be similarly high for wind, water or animal dispersed plants. But the landscape structure and human activity influence this spreading, with higher rates found in densely inhabited or particularly economically active regions (Williamson et al. 2005).

A solution to resolve this discrepancy between estimated dispersal distances and observed migration rates is to consider that dispersal vectors indicated by seed morphology mainly explain the short dispersal distances, with the rare events responsible for LDD relying on other vectors (Cain et al. 1998; Higgins et al. 2003). For example, 78% of the plants that arrived on Surtsey island (Iceland) were transported by water when only one quarter of those taxa were morphologically adapted for water dispersal (Higgins et al. 2003). Birds can transport seeds in mud sticking to their feet (Carlquist 1967), ingest some anemochorous seeds (Wilkinson 1997) or use them to build their nest (*Salix sp*. or *Clematis vitalba*; Müller-Schneider 1983; Dean et al. 1990). Seed plumes or pappus are not only very efficient for wind dispersal (anemochory), but also for fixing on animal fur (Fischer et al. 1996; Couvreur et al. 2004). Finally, humans also are efficient involuntary dispersal vectors nowadays (e.g. Hodkinson and Thompson 1997), but were also vectors during the post-glacial recolonisation, like for *Corylus avellana* or agricultural weeds (Braun-Blanquet 1970; Lang 1994; Clark et al. 1998).

Taking into account the influence of LDD on plant migration in a better manner can possibly be achieved by improving the models fitted on dispersal observations (Kot et al. 1996; Clark et al. 1998; Higgins and Richardson 1999). These improved models would help to propose dispersal distance values for the remaining 1% of the seeds (Tab. 1). But up to now, the necessary values for this improvement are missing for most species and dispersal modes, and hence we cannot propose realistic values for our dispersal distance types. Yet, even though this improvement in LDD estimation would be achieved, it could explain only a part of LDD, as the randomness of unconventional dispersal vectors cannot be standardised for all species. The importance of these accidental dispersions is not known in nature. It may be an important factor for colonising large, new areas (Higgins et al. 2003) or disturbed areas (Bergelson et al. 1993; Williamson et al. 2005), but it is certainly less frequent in closed,

natural vegetation. Takahashi and Kamitani (2004) observed the colonisation of native herbaceous species in an artificial pine forest. They found that the distances dispersed by species using various dispersal vectors were similar to what we proposed for our dispersal types (Tab. 1), thus indicating that the unconventional dispersal vectors were certainly not predominant.

Conclusions

Although the data compiled in this paper are certainly incomplete, they are the most comprehensive data set currently available for the Central European flora. Our method for estimating dispersal distances based on dispersal types is less precise than the calculation of species-specific dispersal models. On the other hand, our typology can be applied to almost all European plant species, which is not the case of a species specific model. As discussed above, our typology is able to represent a large fraction of the interspecies variation in dispersal distances as long as long-distance dispersal is ignored.

Future research on dispersal mechanisms as well as the inclusion of our estimates in species distribution models will show whether the use of this typology leads to predicted migration rates that are close to the observed ones. If differences prove to be important for some of the dispersal types, our typology could be improved by adjusting the corresponding dispersal distances. Alternatively, if observed migration rates are consistently underestimated by the use of our typology, this would suggest that long-distance dispersal is much more important for long-term plant displacements than the dispersal modes presented here. Our typology is therefore certainly not the final one, but an important basis for improving predictive models of species distributions.

Supplementary Material

Appendix 1. Literature data on seed dispersal distances. The examples are mainly from the Swiss flora, except when data were insufficient for certain dispersal modes. Some supplementary species were thus added, mostly from temperate regions. Asterisks indicate values that were considered to represent long-distance dispersal and were therefore excluded from data analysis.

Acknowledgements

This research has been partly supported by the Federal Office for the Environment. We are grateful to Christophe Randin and Antoine Guisan for useful advice and discussion. We thank S. Güsewell, J. Kollmann and an anonymous reviewer for their useful comments on an earlier version of the manuscript.

Résumé (French abstract)

La capacité des plantes à disperser est un facteur important à leur survie dans un paysage fragmenté ou sous l'influence des changements climatiques. Il est donc important de pouvoir tenir compte des distances de dispersion dans les modèles de répartition des espèces, mais les valeurs existantes, mesurées ou calculées, sont rares. Nous proposons donc une approche simple permettant d'estimer ces distances pour l'ensemble d'une flore régionale. Nous avons recherché dans la littérature les données disponibles pour la flore des régions tempérées (avant tout pour les espèces suisses) et associé les distances de dispersion trouvées avec le mode de dispersion et des traits biologiques. Sept types de dispersion ont pu être identifiés sur la base de ces informations, chaque type regroupant des espèces avec des distances de dispersion proches. Les distances à l'intérieur desquelles 50 % et 99 % des graines sont dispersées ont été estimées sur la base du 80^{e} percentile des valeurs disponibles au sein de chaque type. Ces distances varient d'un facteur 5000 entre les sept types de dispersion, alors que les valeurs à disposition pour chaque type ne dépassent généralement pas un facteur de 50. Nos types de dispersion conservent donc une large part de la variation existante dans la dispersion des graines. L'attribution d'une espèce à un type de dispersion ne nécessite que des informations couramment disponibles, comme le vecteur de dispersion (vent, animaux, …), le mode précis de la dispersion (dyszoochorie, épizoochorie, …) et des traits biologiques influençant la dispersion (hauteur de la plante, habitat, …). Cette typologie pourrait être étendue à d'autres régions et permet d'inclure la dispersion des graines dans les modèles de répartition des espèces.

References

Aeschimann D., Heitz C., Palese R., Perret P. et Moser D.M. 1996. Index synonymique de la Flore de Suisse et territoires limitrophes (ISFS). CRSF, Genève.

Bergelson J., Newman J.A. and Floresroux M.E. 1993. Rates of weed spread in spatially heterogeneous environments. *Ecology* **74**: 999-1011.

Bill H.-C., Poschlod P., Reich M. and Plachter H. 1999. Experiments and observations on seed dispersal by running water in Alpine floodplain. *Bull. Geobot. Inst. ETH* **65**: 13-28.

Bonn S. 2004. Dispersal of plants in the Central European landscapes – dispersal processes and assessment of dispersal potential exemplified for endozoochory. PhD thesis, Universität Regensburg.

Bonn S. und Poschlod P. 1998. Ausbreitungsbiologie der Pflanzen Mitteleuropas. Quelle and Meyer, UTB, Wiesbaden.

Bonn S., Poschlod P. and Tackenberg O. 2000. "Diasporus" – a database for diaspore dispersal. Concept and applications in case studies for risk assessment. *Z. Ökologie u. Naturschutz* **9**: 85-97.

Braun-Blanquet J. 1970. Associations messicoles du Languedoc. Leur origine, leur âge. *Melhoramento* **22**: 55-75.

Bullock J.M. and Clarke R.T. 2000. Long distance seed dispersal by wind: measuring and modelling the tail of the curve. *Oecologia* **124**: 506-521.

Cain M.L., Damman H. and Muir A. 1998. Seed dispersal and the Holocene migration of woodland herbs. *Ecol. Monogr.* **68**: 325-347.

Carlquist S. 1967. The biota of long distance dispersal. V. *Plant dispersal to Pacific islands. Bull. Torrey Bot. Club* **94**: 129-162.

Cherix D. 1981. Contribution à la biologie et à l'écologie de *Formica lugubris* Zett. (*Hymenoptera, Formicidae*). Le problème des super-colonies. Thèse de doctorat, Université de Lausanne.

Clark J.S., Fastie C., Hurtt G., Jackson S.T., Johnson C., King G.A., Lewis M., Lynch J., Pacala S., Prentice C., Schupp E.W., Webb T. and Wyckoff P. 1998. Reid's paradox of rapid plant migration (dispersal theory and interpretation of paleoecological records). *Bioscience* **48**:13-24.

Couvreur M., Christiaen B., Verheyen K. and Hermy M. 2004. Large herbivores as mobile links between isolated nature reserves through adhesive seed dispersal. *Appl. Veg. Sci.* **7**: 229-236.

Darwin C.R. 1859. The origin of species. John Murray, London.

Davis A.J., Jenkinson L.S., Lawton J.H., Shorrocks B. and Wood S. 1998. Making mistakes when predicting shifts in species range in response to global warming. *Nature* **391**: 783-786.

Dean W.R.J., Milton S.J. and Siegfried W.R. 1990. Dispersal of seeds as nest material by birds in semiarid karoo shrubland. *Ecology* **71**: 1299-1306.

Dirnbock T., Dullinger S. and Grabherr G. 2003. A regional impact assessment of climate and land-use change on alpine vegetation. *J. Biogeogr.* **30**: 401-417.

Fischer S.F., Poschlod P. and Beinlich B. 1996. Experimental studies on the dispersal of plants and animals on sheep in calcareous grassland. *J. Appl. Ecol.* **33**: 1206-1222.

Greene D.F. and Calogeropoulos C. 2002. Measuring and modelling seed dispersal of terrestrial plants. In: Bullock J.M., Kenward R.E. and Hails R.S. (eds) Dispersal ecology. The 42nd Symposium of the British Ecological Society held at the University of Reading 2-5 April 2001. Blackwell, Oxford, 3-23.

Greene D.F. and Johnson E.A. 1997. Secondary dispersal of tree seeds on snow. *J. Ecol.* **85**: 329-340.

Guisan A. and Theurillat J.-P. 2000. Assessing alpine plant vulnerability to climate change: a modeling perspective. *Integrated Assessment* **1**: 307-320.

Hickey J.R., Flynn R.W., Buskirk S.W., Gerow K.G. and Willson M.F. 1999. An evaluation of a mammalian predator, *Martes americana*, as a disperser of seeds. *Oikos* **87**: 499-508.

Higgins S.I., Nathan R. and Cain M.L. 2003. Are long distance dispersal events in plants usually caused by nonstandard means of dispersal? *Ecology* **84**: 1945-1956.

Higgins S.I. and Richardson D.M. 1999. Predicting plant migration rates in a changing world: the role of long-distance dispersal. *Am. Nat.* **153**: 464-475.

Hodkinson D.J. and Thompson K. 1997. Plant dispersal: The role of man. *J. Appl. Ecol.* **34**: 1484-1496.

Janzen D.H. 1984. Dispersal of small seeds by big herbivores: Foliage is the fruit. *Am. Nat.* **123**: 338-353.

Jongejans E. and Telenius A. 2001. Field experiments on seed dispersal by wind in ten umbelliferous species (*Apiaceae*). *Plant Ecology* **152**: 67-78

Kiviniemi K. and Telenius A. 1998. Experiments on adhesive dispersal by wood mouse: seed shadows and dispersal distances of 13 plant species from cultivated areas in southern Sweden. *Ecography* **21**: 108-116.

Kollmann J. and Schill H.P. 1996. Spatial patterns of dispersal, seed predation and germination during colonization of abandoned grassland by *Quercus petraea* and *Corylus avellana*. *Vegetatio* **125**: 193-205.

Kot M., Lewis M.A. and van den Driessche P. 1996. Dispersal data and the spread of invading organisms. *Ecology* **77**: 2027-2042.

Lang G. 1994. Quartäre Vegetationsgeschichte Europas. Methoden und Ergebnisse. Fischer Verlag, Jena.

Magri D., Vendramin G.G., Comps B., Dupanloup I., Geburek T., Gomory D., Latalowa M., Litt T., Paule L., Roure J.M., Tantau I., van der Knaap W.O., Petit R.J. and de Beaulieu J.L. 2006. A new scenario for

the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytol.* **171**: 199-221.

McCanny S.J. and Cavers P.B. 1988. Spread of proso millet (*Panicum miliaceum* L.) in Ontario, Canada. II. Dispersal by combines. *Weed Res.* **28**: 67-72.

Malcolm J.R., Markham A., Neilson R.P. and Garaci M. 2002. Estimated migration rates under scenarios of global climate change. *J. Biogeogr.* **29**: 835-849.

Mattes H. 1982. Die Lebensgemeinschaft von Tannenhäher, *Nucifraga caryocatactes* (L.), und Arve, *Pinus cembra* L.. *Ber. Eidg. Anst. Forstl. Versuchswes.* **241**: 1-74.

Mouissie A.M., Lengkeek W. and van Diggelen R. 2005a. Estimating adhesive seed-dispersal distances: field experiments and correlated random walks. *Funct. Ecol.* **19**: 478-486.

Mouissie A.M., van der Veen C.E.J., Veen G.F. and van Diggelen R. 2005b. Ecological correlates of seed survival after ingestion by fallow deer. *Funct. Ecol.* **19**: 284-290.

Müller-Schneider P. 1949. Unsere Vögel als Samenverbreiter. *Orn. Beob.* **46**: 120-123.

Müller-Schneider P. 1963. Neue Beobachtungen über Samenverbreitung durch Ameisen. *Ber. Schweiz. Bot. Ges.* **73**: 153-160.

Müller-Schneider P. 1983. Verbreitungsbiologie (Diasporologie) der Blütenpflanzen. 3. *Aufl. Veröff. Geobot. Inst. ETH Stiftung Rübel Zürich* **61**: 1-226.

Müller-Schneider P. 1986. Verbreitungsbiologie der Blütenpflanzen Graubündens. *Veröff. Geobot. Inst. ETH Stiftung Rübel Zürich* **85**: 1-263.

Nathan R., Sapir N., Trakhtenbrot A., Katul G.G., Bohrer G., Otte M., Avissar R., Soons M.B., Horn H.S., Wikelski M. and Levin S.A. 2005. Long-distance biological transport processes through the air: can nature's complexity be unfolded in silico? *Diversity Distrib.* **11**: 131-137.

Pakeman R.J. 2001. Plant migration rates and seed dispersal mechanisms. *J. Biogeogr.* **28**: 795-800.

Pearson R.G. 2006. Climate change and the migration capacity of species. *Trends Ecol. Evol.* **21**: 111-113.

Pitelka L.F., Gardner R.H., Ash J. Berry S., Gitay H., Noble I.R., Saunders A., Bradshaw R.H.W., Brubaker L., Clark J.S., Davis M.B., Sugita S., Dyer J.M., Hengeveld R., Hope G., Huntley B., King G.A., Lavorel S., Mack R.N., Malanson G.P., McGlone M., Prentice I.C. and Rejmanek M. 1997. Plant migration and climate change. *Am. Sci.* **85**: 464-473.

Pyšek P. and Hulme P.E. 2005. Spatio-temporal dynamics of plant invasions: linking pattern to process. *Ecoscience* **12**: 302-315.

Reid C. 1899. The origin of the British flora. Dulau, London.

Ridley H.N. 1930. The dispersal of plants throughout the world. Reeve, Ashford.

Römermann C., Tackenberg O. and Poschlod P. 2005. How to predict attachment potential of seeds to sheep and cattle coat from simple morphological seed traits. *Oikos* **110**: 219-230.

Ronce O. 2001. Understanding plant dispersal and migration. *Trends Ecol. Evol.* **16**: 663.

Schmidt W. 1918. Die Verbreitung von Samen und Blütenstaub durch die Luftbewegung. Oesterr. *Bot. Z.* **67**: 313-328.

Schneider S. 1935. Untersuchungen und Samenschleudermechanismen verschiedener Rhoeadales. *Jahrb. Wiss. Botanik* **81**: 663-704.

Skellam J.G. 1951. Random dispersal in theoretical populations. *Biometrika* **38**: 196-218.

Soons M.B. and Ozinga W.A. 2005. How important is long-distance seed dispersal for the regional survival of plant species? *Diversity Distrib.* **11**: 165-172.

Stewart J.R. and Lister A.M. 2001. Cryptic northern refugia and the origins of the modern biota. Trends. *Ecol. Evol.* **16**: 608-613.

Stöcklin J. and Bäumler E. 1996. Seed rain, seedling establishment and clonal growth strategies on a glacier foreland. *J. Veg. Sci.* **7**: 45-56.

Tackenberg O. 2001. Methoden zur Bewertung gradueller Unterschiede des Ausbreitungspotentials von Pflanzenarten. PhD thesis, Philipps-Universität Marburg.

Tackenberg O., Poschlod P. and Bonn S. 2003. Assessment of wind dispersal potential in plant species. *Ecol. Monogr.* **73**: 191-205.

Takahashi K. and Kamitani T. 2004. Effect of dispersal capacity on forest plant migration at landscape scale. *J. Ecol.* **92**: 778-785.

Thuiller W., Lavorel S., Araújo M.B., Sykes M.T. and Prentice I.C. 2005. Climate change threats to plant diversity in Europe. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 8245-8250.

Tinner U. and Schumacher H. 2004. Flora auf Bahnhöfen der Nordostschweiz. *Bot. Helv.* **114**: 109-125.

Vellend M., Myers J.A., Gardescu S. and Marks P.L. 2003. Dispersal of Trillium seeds by deer: implications for long-distance migration of forest herbs. *Ecology* **84**: 1067-1072.

Wilkinson D.M. 1997. Plant colonization: are wind dispersed seeds really dispersed by birds at large spatial and temporal scales? *J. Biogeogr.* **24**: 61-65.

Williamson M., Pyšek P., Jarošík V. and Prach K. 2005. On the rates and patterns of spread of alien plants in the Czech Republic, Britain, and Ireland. *Ecoscience* **12**: 424-433.

Xiao Z., Zhang Z. and Wang Y. 2004. Dispersal and germination of big and small nuts of *Quercus serrata* in a subtropical broad-leaved evergreen forest. *Forest Ecol. Manage.* **195**: 141–150.

Supplementary Material / Supporting Information

Appendix 1

Literature data on seed dispersal distances. The examples are mainly from the Swiss flora, except when data were insufficient for certain dispersal modes. Some supplementary species were thus added, mostly from temperate regions. Asterisks indicate values that were considered to represent long-distance dispersal and were therefore excluded from data analysis.

Species	Distance [m] * denotes long-distance dispersal	Type of Measurement	Reference
Autochory			
Ballochory			
Cardamine amara	1.4	Maximum	Schneider 1935
Cardamine impatiens	2	Maximum	Schneider 1935
Cardamine pratensis	2.4	Maximum	Schneider 1935
Cardamine resedifolia	1.04	Maximum	Schneider 1935
Cardamine resedifolia	< 1	Maximum	Stöcklin & Bäumler 1996
Geranium maculatum	1.4 / 6.2	Mode and maximum	Stamp & Lucas 1983 and unpubl. in Willson 1993
Geranium maculatum	4.6 / 5.6	Mode and maximum	Stamp & Lucas 1983 and unpubl. in Willson 1993
Geranium molle	1.78 / 2.8	Mean and maximum	Stamp & Lucas 1983 in Cain et al. 1998
Geranium robertianum	6	Maximum	Ridley 1930 in Müller-Schneider 1983
Geranium rotundifolium	1.8	Maximum	Müller-Schneider 1933 in Müller-Schneider 1983
Geranium sylvaticum	2.7	Maximum	Müller-Schneider 1983
Impatiens parviflora	3.4	Maximum	Schneider 1935
Lathraea clandestina	4	Maximum	Guttenberg 1926 in Müller-Schneider 1983
Lathyrus vernus	1.5	Maximum	Müller-Schneider 1986
Mercurialis annua	2.9	Maximum	Müller-Schneider 1983
Mercurialis perennis	4	Maximum	Ridley 1930 in Müller-Schneider 1983
Mercurialis perennis	0.89	Maximum	Müller-Schneider 1986
Oxalis acetosella	2.3	Maximum	Moor 1940 in Müller-Schneider 1983
Oxalis acetosella	5	Maximum	Berg 2000
Viola arvensis	2.4	Maximum	Stapf 1887 in Müller-Schneider 1983
Viola canina	4.7	Maximum	Ulbrich 1928 in Müller-Schneider 1983
Viola riviniana	4.6	Maximum	Ulbrich 1928 in Müller-Schneider 1983
Viola stricta	0.5 / 3.4	Mode and maximum	Stamp & Lucas 1983 in Willson 1993
Anemochory			
Cystometeorochory			
Calluna vulgaris	1 / 80*	90% of the seeds and maximum	Bullock & Clarke 2000
Calluna vulgaris	0.35 / 2	Calculated median and 99-percentile	Soons & Ozinga 2005
Erica cinerea	1 / 80*	90% of the seeds and maximum	Bullock & Clarke 2000
Sanguisorba minor	0.03 / 0.17	Calculated median and 99-percentile	Soons & Ozinga 2005
Orchidaceae			
Cephalanthera damasonium	0.95 / 14.7	Calculated median and 99-percentile	Soons & Ozinga 2005
Trichometeorochory			
Little efficient			
Anthoxanthum odoratum	0.3 / 2	Mode and maximum	Antonovics & Ellstrand 1985 in Willson 1993

Carex frigida	< 1	Maximum	Stöcklin & Bäumler 1996
Carlina vulgaris	0.22 / 1.7	Calculated median and 99-percentile	Soons & Ozinga 2005
Carlina vulgaris	1.47	Maximum with a 16.4 km/h wind	Sheldom & Burrows 1973 in Cain et al. 1998
Crepis paludosa	0.31 / 2.2	Calculated median and 99-percentile	Soons & Ozinga 2005
Geum reptans	4	Maximum	Stöcklin & Bäumler 1996
Hieracium aurantiacum	0.1 / 1.9	Mode an maximum	Stergios 1976 in Willson 1993
Hieracium murorum aggr.	0.27 / 1.9	Calculated median and 99-percentile	Soons & Ozinga 2005
Hieracium murorum aggr.	10	Maximum	Stöcklin & Bäumler 1996
Hieracium pilosella	0.21 / 1.7	Calculated median and 99-percentile	Soons & Ozinga 2005
Hieracium sabaudum	1.5 / 15.7	Calculated median and 99-percentile	Soons & Ozinga 2005
Leontodon autumnalis	0.12 / 0.81	Calculated median and 99-percentile	Soons & Ozinga 2005
Picris hieracioides	0.40 / 3.54	Calculated median and 99-percentile	Soons & Ozinga 2005
Pulsatilla sp.	80*	Maximum	Hegi 1906-1938 in Müller-Schneider 1986
Senecio jacobea	0.49 / 4.1	Calculated median and 99-percentile	Soons & Ozinga 2005
Senecio jacobea	14 / 36	Mode and maximum	McEvoy & Cox 1987 in Willson 1993
Senecio vulgaris	0.27 / 3.4	Calculated median and 99-percentile	Soons & Ozinga 2005
Senecio vulgaris	0.34 / 2	Mean and maximum	Bergelson et al. 1993
Tragopogon pratensis	0.41 / 3.4	Calculated median and 99-percentile	Soons & Ozinga 2005
Forest plants			
Mycelis muralis	0.99 / 14.3	Calculated median and 99-percentile	Soons & Ozinga 2005
Solidago virgaurea	0.58 / 5.6	Calculated median and 99-percentile	Soons & Ozinga 2005
Highly efficient			
Adenostyles leucophylla	85	Maximum	Stöcklin & Bäumler 1996
Carduus nutans	0.83 / 9.6	Calculated median and 99-percentile	Soons & Ozinga 2005
Carduus nutans	10 / 40-100	Mode and maximum of different measures	Smith & Kok 1984 in Willson 1993
Cirsium arvense	2.0 / 53.4	Calculated median and 99-percentile	Soons & Ozinga 2005
Cirsium arvense	11.4	Maximum with a 16.4 km/h wind	Sheldom & Burrows 1973 in Cain et al. 1998
Cirsium spinosissimum	30	Maximum	Stöcklin & Bäumler 1996
Cirsium vulgare	1.8 / 31.6	Calculated median and 99-percentile	Soons & Ozinga 2005
Cirsium vulgare	1 / 32	Mean and maximum	Klinkhammer et al. 1988 in Cain et al. 1998
Clematis sp.	100	Maximum	Müller-Schneider 1986
Clematis vitalba	10.2 / 100	Calculated median and 99-percentile	Soons & Ozinga 2005
Epilobium angustifolium	7.48 / 2112	Calculated median and 99-percentile	Soons & Ozinga 2005
Epilobium ciliatum	3.65 / 179	Calculated median and 99-percentile	Soons & Ozinga 2005
Epilobium fleischeri	50	Maximum	Stöcklin & Bäumler 1996
Epilobium hirsutum	4 / 136	Calculated median and 99-percentile	Soons & Ozinga 2005
Epilobium montanum	1.6 / 49.6	Calculated median and 99-percentile	Soons & Ozinga 2005
Erigeron acer s.l.	75	Maximum	Stöcklin & Bäumler 1996
Erigeron annuus	1.6 / 35.4	Calculated median and 99-percentile	Soons & Ozinga 2005
Eriophorum angustifolium	1.1 / 24.2	Calculated median and 99-percentile	Soons & Ozinga 2005
Eupatorium cannabinum	1.6 / 23.7	Calculated median and 99-percentile	Soons & Ozinga 2005
Hieracium staticifolium	75	Maximum	Stöcklin & Bäumler 1996
Myricaria germanica	100	Maximum	Stöcklin & Bäumler 1996
Phragmites australis	13.9 / 1714	Calculated median and 99-percentile	Soons & Ozinga 2005
Poa nemoralis	50	Maximum	Stöcklin & Bäumler 1996
Salix sp.	100	Maximum	Stöcklin & Bäumler 1996
Solidago gigantea	4.2 / 136	Calculated median and 99-percentile	Soons & Ozinga 2005
Solidago virgaurea subsp. minuta	4	Maximum	Stöcklin & Bäumler 1996
Taraxacum officinale	0.22 / 2.2	Calculated median and 99-percentile	Soons & Ozinga 2005
Taraxacum officinale	50	Maximum	Stöcklin & Bäumler 1996
Tussilago farfara	20	Maximum	Stöcklin & Bäumler 1996
Tussilago farfara	10 / > 4000*	Mode and maximum	Bakker 1961 in Willson 1993

Typha angustifolia	11.3 / 2194	Calculated median and 99-percentile	Soons & Ozinga 2005
Typha latifolia	14.7 / 3673	Calculated median and 99-percentile	Soons & Ozinga 2005
Pterometeorochory			
Herbs			
Agrostis rupestris	< 1	Maximum	Stöcklin & Bäumler 1996
Angelica sylvestris	2.29	Median with a 5.3 m/s wind	Jongejans & Telenius 2001
Angelica sylvestris	0.31 / 1.91	Calculated median and 99-percentile	Soons & Ozinga 2005
Bromus inermis	1.72 / 7	Mean and maximum	Hume & Archbold 1986 in Cain et al. 1998
Bromus sterilis	20	Maximum	Howard et al. 1992 in Bullock & Clarke 2000
Heracleum sphondylium	0.38 / 2.11	Calculated median and 99-percentile	Soons & Ozinga 2005
Heracleum sphondylium	3.12	Median with a 4.7 m/s wind	Jongejans & Telenius 2001
Laserpitium latifolium	1.9	Median with a 4.3 m/s wind	Jongejans & Telenius 2001
Oxyria digyna	1	Maximum	Stöcklin & Bäumler 1996
Pastinaca sativa	3.05	Median with a 4.7 m/s wind	Jongejans & Telenius 2001
Peucedanum palustre	0.25 / 1.49	Calculated median and 99-percentile	Soons & Ozinga 2005
Peucedanum palustre	1.31	Median with a 3.4 m/s wind	Jongejans & Telenius 2001
Rumex acetosa	0.18 / 0.99	Calculated median and 99-percentile	Soons & Ozinga 2005
Rumex scutatus	12	Maximum	Stöcklin & Bäumler 1996
Scabiosa columbaria	1.9	Maximum	Verkaar et al. 1983
Selinum carvifolia	0.1 / 0.5	Calculated median and 99-percentile	Soons & Ozinga 2005
Selinum carvifolia	0.79	Median with a wind of 2.6 m/s	Jongejans & Telenius 2001
Trifolium badium	10	Maximum	Stöcklin & Bäumler 1996
Trifolium pallescens	6	Maximum	Stöcklin & Bäumler 1996
Trees			
Abies alba	7000*	Maximum	Bouget & Davy de Virville 1926 in Müller-Schneider 1983
Acer pseudoplatanus	5000*	Maximum	Braun-Blanquet 1913 in Müller-Schneider 1986
Acer pseudoplatanus	400-500*	Maximum	Firbas 1935 in Müller-Schneider 1986
Acer rubrum	83 / 314 / 11'371*	Calculated median, 99-percentile et maximum	Higgins et al. 2003
Alnus viridis	70	Maximum	Stöcklin & Bäumler 1996
Betula sp.	40 / 100	Limit for the majority and maximum	Greene & Calogeropoulos 2002
Carpinus betulus	130	Maximum	Müller-Schneider 1986
Fraxinus excelsior	725*	Maximum	Geiger 1960 in Müller-Schneider 1986
Fraxinus sp.	40 / 100	Limit for the majority and maximum	Greene & Calogeropoulos 2002
Larix decidua	15	Maximum	Stöcklin & Bäumler 1996
Picea abies	1500* / 800*	Maximum horizontal and vertical dispersion	Braun-Blanquet 1913 in Müller-Schneider 1986
Picea glauca	475*	Maximum	Greene & Johnson 1995 in Cain et al. 1998
Pinus sp.	40 / 100	Limit for the majority and maximum	Greene & Calogeropoulos 2002
Pinus sylvestris	2000*	Maximum	Firbas 1935 in Müller-Schneider 1983
Tilia platyphyllos	80	Maximum observed	Müller-Schneider 1986
Chamaechory			
Betula alleghaniensis	85 / 200	Optimum and maximum	Greene & Johnson 1997
Boleochory			
Short species			
Achillea erba-rotta subsp. moschata	< 1	Maximum	Stöcklin & Bäumler 1996
Achillea millefolium	0.07 / 0.39	Calculated median and 99-percentile	Soons & Ozinga 2005
Achillea nana	4*	Maximum	Stöcklin & Bäumler 1996
Arabis alpina	< 1	Maximum	Stöcklin & Bäumler 1996
Arabis hirsuta	0.09 / 0.47	Calculated median and 99-percentile	Soons & Ozinga 2005
Arenaria serpyllifolia	0.03 / 0.13	Calculated median and 99-percentile	Soons & Ozinga 2005
Bellis perennis	0.02 / 0.09	Calculated median and 99-percentile	Soons & Ozinga 2005

Campanula rotundifolia	0.07 / 0.35	Calculated median and 99-percentile	Soons & Ozinga 2005
Capsella bursa-pastoris	0.05 / 0.23	Calculated median and 99-percentile	Soons & Ozinga 2005
Cerastium fontanum subsp. vulgare	0.03 / 0.16	Calculated median and 99-percentile	Soons & Ozinga 2005
Cerastium arvense	< 1	Maximum	Stöcklin & Bäumler 1996
Cerastium pedunculatum	< 1	Maximum	Stöcklin & Bäumler 1996
Eranthis hiemalis	0.32 / 1.23	Median and maximum in natural wind	Emig et al. 1999
Gentiana germanica	1.2	Maximum	Verkaar et al. 1983
Linaria alpina	12*	Maximum	Stöcklin & Bäumler 1998
Linum catharticum	0.02 / 0.13	Calculated median and 99-percentile	Soons & Ozinga 2005
Linum catharticum	0.41 / 0.63	Maximum in dense vegetation or open micro-sites	Verkaar et al. 1983
Primula veris	0.03 / 0.12	Calculated median and 99-percentile	Soons & Ozinga 2005
Sagina saginoides	10*	Maximum	Stöcklin & Bäumler 1996
Saxifraga sp.	40*	Maximum	Stöcklin & Bäumler 1996
Saxifraga tridactylites	0.02 / 0.11	Calculated median and 99-percentile	Soons & Ozinga 2005
Sempervivum sp.	< 1	Maximum	Stöcklin & Bäumler 1996
Silene rupestris	10*	Maximum	Stöcklin & Bäumler 1996
Tall species			
Aquilegia vulgaris	0.07 / 0.41	Calculated median and 99-percentile	Soons & Ozinga 2005
Campanula trachelium	0.25 / 1.45	Calculated median and 99-percentile	Soons & Ozinga 2005
Daucus carotta	0.15 / 0.93	Calculated median and 99-percentile	Soons & Ozinga 2005
Oenothera biennis	0.15 / 0.98	Calculated median and 99-percentile	Soons & Ozinga 2005
Oenothera biennis	1.6 / 5	Mode an maximum	Platt & Weis 1977 in Willson 1993
Papaver argemone	0.6 / 3.1	Mode an maximum	Salisbury 1942 in Willson 1993
Papaver dubium	0.9 / 2.1	Mode an maximum	Salisbury 1942 in Willson 1993
Papaver rhoeas	0.1 / 0.5	Calculated median and 99-percentile	Soons & Ozinga 2005
Papaver somniferum	2	Maximum with limited wind	Müller-Schneider 1983
Rhododendron ferrugineum	25*	Maximum	Stöcklin & Bäumler 1996
Silene pratensis	0.1 / 0.57	Calculated median and 99-percentile	Soons & Ozinga 2005
Verbascum thapsus	0.15 / 0.87	Calculated median and 99-percentile	Soons & Ozinga 2005
Hydrochory			
Ombrochory			
Eranthis hiemalis	0.4	Maximum	Müller-Schneider 1936 in Müller-Schneider 1983
Eranthis hiemalis	0.17 / 1.17	Median and maximum dispersion under the rain	Emig et al. 1999
Thlaspi perfoliatum	0.8	Maximum	Müller-Schneider 1936 in Müller-Schneider 1983
Zoochory			
Dyszoochory			
Small animals			
Fagus sylvatica	4.13 / 13	Mean and maximum by rodents *(Clethrionomys sp.* and *Apodemus sp.)*	Jensen 1985 in Cain et al. 1998
Helianthus annuus	20	Mean by nuthatch (*Sitta europaea*)	Müller-Schneider 1949
Juglans nigra	15 / 38.1 / 151	Minimum, mean and maximum by fox squirrels (*Sciurus niger*)	Stapanian & Smith 1978, 1986
Picea abies	60	By great spotted woodpecker (*Dendrocopos major*) to open the cones	Müller-Schneider 1983
Pinus strobus	15 / 30	70 % and maximum by rodents (*Peromyscus sp.* and *Clethrionomys sp.*)	Abbott & Quink 1970
Pinus jeffreyi	21 / 69	Mean and maximim by rodents (chipmunk, *Tamias sp.*)	Vander Wall 1993 in Cain et al. 1998
Pinus sp.	1800*	Maximal dispersion by red squirrels (*Sciurus vulgaris orientis*)	Hayashida 1988
Quercus serrata	10 / 28.5	80% of the seeds and maximum by small rodents (mice and rats)	Xiao et al. 2004
Quercus macrocarpa	10 / 52	Mean and maximum by fox squirrels (*Sciurus niger*)	Stapanian & Smith 1986
Quercus petraea	18	Maximum by rodents (*Apodemus sp.*)	Kollmann & Schill 1996
Quercus sp.	15.3 / 34	Mean and maximum by rodents (*Apodemus sp.* and *Clethrionomys sp.*)	Jensen & Nielsen 1986
Large animals			
Corylus avellana	15000*	Maximum dispersion by nutcrackers (*Nucifraga caryocatactes*)	Mattes 1982

Fagus grandifolia	4000*	Maximum dispersion by blue jays (*Cianocitta cristata*)	Johnsonn & Adkinson 1985 in Clark et al. 1998
Fagus sylvatica	32	Dispersion by jay (*Garrulus glandarius*)	Müller-Schneider 1949
Juglans regia	200	By carrion crow (*Corvus corone*) to break the nut	Müller-Schneider 1983
Pinus albicaulis	100 / 3500	Mean and maximum by birds	Hutchins & Lanner 1982 in Cain et al. 1998
Pinus cembra	12'000*	Maximum dispersion by nutcrackers (*Nucifraga caryocatactes*)	Sutter & Ammann 1953 in Müller-Schneider 1986
Quercus palustris	1100 / 1900	Mean and maximum dispersion by birds	Darley-Hill & Johnson 1981 in Cain et al. 1998
Quercus sp.	4000*	Maximum dispersion by jay (*Garrulus glandarius*)	Müller-Schneider 1983
Quercus petraea	300	Maximum by jay (*Garrulus glandarius*)	Kollmann & Schill 1996
Myrmecochory			
Allium ursinum	1.52-4.61	Different observations with *Formica rufa*	Müller-Schneider 1971
Allium ursinum	0.95	One observation with *Formica cinerea*	Müller-Schneider 1971
Asarum canadense	1.54 / 35	Mean and maximum	Cain et al. 1998
Carex pilulifera	0.75 / 1.4	Mean and maximum dispersed by *Myrmica ruginodis*	Kjellsson 1985 in Ness et al. 2004
Chelidonium majus	80*	Maximum	Senander 1906 in Bonn & Poschold 1998
Daphne striata	6.38	One observation with *Formica lugubris*	Müller-Schneider 1963
Euphorbia characias	2.1 / 4.6	Mean and maximum dispersed by *Aphaenogaster senilis*	Gomez & Espadaler 1998 in Ness et al. 2004
Euphorbia characias	2.1 / 9.4	Mean and maximum dispersed by *Messor barbarus*	Gomez & Espadaler 1998 in Ness et al. 2004
Euphorbia characias	0.79 / 1.6	Mean and maximum dispersed by *Tapinoma nigerrimum*	Gomez & Espadaler 1998 in Ness et al. 2004
Melica nutans	70*	Maximum	Senander 1906 in Bonn & Poschold 1998
Mercurialis annua	3.4 / 14	Mean and maximum dispersed by *Messor structor*	Lisci & Pacini 1997 in Ness et al. 2004
Rhamnus alaternus	1 / 5	Mean and maximum	Gomez et al. 2003
Sanguinaria canadensis	17	Maximum	Pudlo et al. 1980 in Cain et al. 1998
Sanguinaria canadensis	2.57 / 6.7	Mean and maximum dispersed by *Formica subsericea*	Ness 2004 in Ness et al. 2004
Viola hirta	70*	Maximum	Senander 1906 in Bonn & Poschold 1998
Viola sp.	0.75 / 1.5	Mean and maximum	Culver & Beattie 1978 in Cain et al. 1998
Various species	70*	Maximum dispersion by *Formica rufa*	Sernander 1906 in Müller-Schneider 1983
Various species	15	Maximum dispersion by *Lasius niger*	Sernander 1906 in Müller-Schneider 1983
Various species	0.96 / 77*	Mean and maximum in world literature	Gomez & Espadaler 1998
Various species	0.64 / 2.7	Mean and maximum in mesic deciduous forest in Japan	Higashi et al. 1989 in Ness et al. 2004
Various species	0.91 / 4.5	Mean and maximum in Oak–Pine temperate woodlands in USA	Gibson 1993 in Ness et al. 2004
Various species	0.53 / 5.2	Mean and maximum in temperate deciduous forest in USA	Kalisz et al. 1999 in Ness et al. 2004
Various species	2.4 / 10	Mean and maximum in temperate deciduous forest in USA	Kalisz et al. 1999 in Ness et al. 2004
Endozoochory			
Prunus avium	> 1000*	Altitudinal shift by fox	Vittoz, unpublished observation
Prunus avium	30 / 100	Mean and maximum by birds	Turcek 1968 in Bonn & Poschold 1998
Prunus serotina	7.1 / 35	Mean and maximum by birds	Smith 1975 in Cain et al. 1998
Rubus idaeus	> 900*	Altitudinal shift by alpine chough (*Pyrrhocorax graculus*)	Müller-Schneider 1983
Parthenocissus quinquefolia	9 / 24	Mean and maximum by birds	Hoppes 1988 in Cain et al. 1998
Phytolacca americana	33	Maximum by birds	Hoppes 1988 in Cain et al. 1998
Trillium grandiflorum	700 / 2500 / 3750*	Median, 99-percentile and maximum by deer (*Odocoileus virginianus*)	Vellend et al. 2003
Vaccinium sp.	500	Median by marten (*Martes americana*)	Hickey et al. 1999
Vitis vulpina	24	Maximum by birds	Hoppes 1988 in Cain et al. 1998
Various species	50 / 180	Maximum and extrem by blackbird (*Turdus merula*)	Müller-Schneider & Lenggenhager 1959 in Bonn & Poschold 1998
Epizoochory			
Small mammals			
Agrimonia eupatoria	11	Maximum by wood mouse (*Apodemum flavicollis*)	Kiviniemi & Telenius 1998
Daucus carotta	17	Maximum by wood mouse (*Apodemum flavicollis*)	Kiviniemi & Telenius 1998
Geum rivale	26	Maximum by wood mouse (*Apodemum flavicollis*)	Kiviniemi & Telenius 1998
Sanicula europaea	15	Maximum by wood mouse (*Apodemum flavicollis*)	Kiviniemi & Telenius 1998
Torilis japonica	15	Maximum by wood mouse (*Apodemum flavicollis*)	Kiviniemi & Telenius 1998
Triglochin palustris	3	Maximum by wood mouse (*Apodemum flavicollis*)	Kiviniemi & Telenius 1998

Various species	12	99-percentile by wood mouse (*Apodemum flavicollis*)	Mouissie et al. 2005a
Large mammals			
Agrimonia eupatoria	932	Maximum by fallow deer (*Dama dama*)	Kiviniemi 1996 in Kiviniemi & Telenius 1998
Agrimonia eupatoria	780	Maximum by cattle	Kivieniemi & Eriksson in Kiviniemi & Telenius 1998
Bidens sp.	109	Mean	Bullock & Primack 1977 in Cain et al. 1998
Geum rivale	660	Maximum by cattle	Kivieniemi & Eriksson in Kiviniemi & Telenius 1998
Jurinea cyanoides	10 / 17	99-percentile and maximum	Eichberg et al. 2005
Triglochin palustris	1242	Maximum by fallow deer (*Dama dama*)	Kiviniemi & Telenius 1998
Various species	380 / 2900	Mode and 99-percentile by sheep	Mouissie et al. 2005a
Various species	65 / 435	Mode and 99-percentile by fallow deer (*Dama dama*)	Mouissie et al. 2005a
Various species	125 / 850	Mode and 99-percentile by cattle	Mouissie et al. 2005a
Anthropochory			
Agochory			
Bromus tectorum	6300	Calculated annual migration rate	Mack 1986 in Malcolm et al. 2002
Veronica filliformis	4700	Calculated annual migration rate	Williamson et al. 2003 in Pyšek & Hulme 2005

References

Abbott H.G. & Qunik T.F. 1970. Ecology of eastern white pine caches made by small forest mammals. *Ecology* **51**: 271-278.

Berg H. 2000. Differential seed dispersal in Oxalis acetosella, a cleistogamous perennial herb. *Acta Oecol.* **21**: 109-118.

Eichberg C., Storm C. and Schwabe A. 2005. Epizoochorous and post-dispersal processes in a rare plant species: Jurinea cyanoides (L.) Rchb. (Asteraceae). *Flora* **200**: 477–489.

Emig W., Hauck I. und Leins P. 1999. Experimentelle Untersuchung zur Samenausbreitung von *Erianthis hyemalis* (L.) Salisb. (Ranunculaceae). *Bull. Geobot. Inst. ETH* **65**:29-41.

Gomez C. and Espadaler X. 1998. Myrmecochorus dispersal distances: a world survey. *J. Biogeogr.* **25**: 573-580.

Gomez C., Pons P. and Bas J.M. 2003. Effects of the Argentine ant Linepithema humile on seed dispersal and seedling emergence of Rhamnus alaternus. *Ecography* **26**: 532–538.

Hayashida M. 1988. Seed dispersal by red squirrels and subsequent establishment of Korean pine. *Forest Ecol. Manage.* **28**: 115-129.

Jensen T.S. & Nielsen O.F. 1986. Rodents as seeds dispersers in a heath oak wood succession. *Oecologia* **70**: 214-221.

Müller-Schneider P. 1971. Beiträge zur Kenntnis der Samenverbreitung durch Ameisen. *Ber. Schweiz. Bot. Ges.* **80**: 289-297.

Ness J.H. 2004. Forest edges and fire ants alter the seed shadow of an ant-dispersed plant. *Oecologia* **138**: 448–454.

Ness J.H., Bronstein J.L., Andersen A.N. and Holland J.N. 2004. Ant body size predicts dispersal distance of ant-adapted seeds: implications of small-ant invasions. *Ecology* **85**: 1244–1250.

Stapanian M.A. and Smith C.C. 1978. A model for seed scatterhoarding: coevolution of fox squirrels and black walnuts. *Ecology* **59**: 884-896.

Stapanian M.A. and Smith C.C. 1986. How fox squirrels influence the invasion of prairies by nut-bearing trees. *J. Mammal.* **67**: 326-332.

Verkaar H.J, Schenkeveld A.J. and van de Klashorst M.P. 1983. The ecology of short-lived forbs in chalk grasslands: Dispersal of seeds. *New Phytol.* **95**: 335-344.

Willson M.F. 1993. Dispersal mode, seed shadows, and colonization patterns. *Vegetatio* **107/108**: 261-280.

2.3

Predicting future distributions of mountain plants under climate change: does dispersal capacity matter?

Robin Engler[1], Christophe F. Randin[1], Pascal Vittoz[1,2], Thomas Czáka[1,2], Mertin Beniston[3], Niklaus E. Zimmermann[4] and Antoine Guisan[1,2]

This manuscript has been published in 2009 *Ecography*, **32**, 34-45.

DOI: 10.1111/j.1600-0587.2009.05789.x

[1] Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland.

[2] Faculty of Geosciences and the Environment, University of Lausanne, Switzerland.

[3] Climate Change and Climate Impacts, (C3i), University of Geneva, Switzerland.

[4] Swiss Federal Research Institute WSL, Land Use Dynamics, Zuercherstrasse 111, CH-8903 Birmensdorf.

The first two authors have contributed equally to this paper.

CONTRIBUTION TO THE PROJECT:
I modified the MIGCLIM program to enable batch simulations. CR prepared the habitat suitability maps and I ran the dynamic simulations. I analysed the results in collaboration with CR and prepared the figures. I wrote the manuscript in collaboration with CR and lead the reviewing process.

Abstract

Many studies have investigated the potential impacts of climate change on the distribution of plant species, but few have attempted to constrain projections through plant dispersal limitations. Instead, most studies published so far have simplified dispersal as either unlimited or null. However, depending on the dispersal capacity of a species, landscape fragmentation, and the rate of climatic change, these assumptions can lead to serious over- or underestimation of the future distribution of plant species.

To quantify the discrepancies between simulations accounting for dispersal or not, we carried out projections of future distribution over the 21st century for 287 mountain plant species in a study area of the Western Swiss Alps. For each species, simulations were run for four dispersal scenarios (unlimited dispersal, no dispersal, realistic dispersal, and realistic dispersal with long-distance dispersal events) and under four climate change scenarios.

Although simulations accounting for realistic dispersal limitations did significantly differ from those considering dispersal as unlimited or null in terms of projected future distribution, the unlimited dispersal simplification did nevertheless provide good approximations for species extinctions under more moderate climate change scenarios. Overall, simulations accounting for dispersal limitations produced, for our mountainous study area, results that were significantly closer to unlimited dispersal than to no dispersal. Finally, analysis of the temporal pattern of species extinctions over the entire 21st century revealed that important species extinctions for our study area might not occur before the 2080-2100 period, due to the possibility of a large number of species shifting their distribution to higher elevation.

Keywords: species distribution model, Generalized Linear Model, cellular automaton, plant species, seed dispersal, Western Swiss Alps, climate change

Introduction

During the last three decades, strong evidence has emerged suggesting that increased atmospheric concentration of greenhouse gases, particularly carbon dioxide and methane linked to human activities, have already begun to modify the global climate (IPCC 2007). On average, global surface temperature increased by 0.61 (±0.18°C) between 1861 and 2000 (Folland et al. 2001).

Evidence from palynology and fossil records have shown that, in the past, plant species have successfully responded to environmental change by (i) remaining within the modified climate by tolerance or adaptation, (ii) migrating to track suitable conditions, or most likely a combination of both (Davis and Shaw 2001, Jump and Peñuelas 2005). However, due to its high magnitude and fast rate, it is uncertain whether plants will be able to withstand the climate change forecasted for the 21st century by tolerance and adaptation solely (Davis and Shaw 2001, Etterson and Shaw 2001, Jump and Peñuelas 2005). The ability of plants to migrate and keep pace with the shift of their suitable habitats is thus likely to be of prime importance for their survival, although this might sometimes require migration rates much higher than those observed from pollen reconstructions (Davis and Shaw 2001, Malcolm et al. 2002). In addition to the rapidity of climate change, a further challenge to plant migration is the human-driven habitat fragmentation that renders the landscape increasingly impassable (Pitelka et al. 1997), and hampers both seed dispersal and gene flow between populations (Davis and Shaw 2001). Finally, natural barriers to dispersal, such as rivers, lakes, forests, or valleys and mountain ranges at a larger scale, represent yet another set of potential obstacles to successful migration.

If a species can neither adapt to the modified environmental conditions nor migrate fast enough, then drastic range reduction or even local extinction – or quasi-extinction – is to be expected. Quasi-extinction indicates that a species could still persist in a few small and scattered locations in the form of relict populations (Eriksson 2000), but these remaining individuals may be very vulnerable to any further disturbances. Reductions in distribution or extinction are particularly expected for species with weak dispersal capacity, long maturation time (i.e., time before the production of propagules), and/or slow growth rate.

A widespread method to assess the impact of future climate change on plants has been the utilization of species distribution models (Guisan and Zimmermann 2000, Guisan and Thuiller 2005). A broad range of projections have already been carried out to forecast future plant species distributions, at large geographical scales with coarse resolution (e.g., at continental scale, Bakkenes et al. 2002, Thuiller et al. 2005), and also at regional or local scales with high resolution data (e.g., Guisan and Theurillat 2000, Dirnböck et al. 2003). Importantly, these models assume that a species is in equilibrium with its environment, and therefore, do not consider possible adaptations or tolerance to climate change. Another limitation of species distribution models, at least in their basic implementation, is that they ignore dispersal restrictions. Thus far, most studies have been based on the assumption of universal dispersal (i.e., a species has unlimited dispersal, its future distribution being the entire area projected by the model; Thomas et al. 2004). While this hypothesis might provide good approximations for plants that are human-dispersed or have high dispersal ability, it is likely overestimating the future distribution of many other species for the previously mentioned reasons, namely the fast rate of the forecasted climate change and habitat fragmentation. As the unlimited dispersal assumption represents an optimistic "best case" scenario, some studies have also provided a "worst case" no-dispersal scenario (e.g., Thuiller et al. 2005, Loarie et al. 2008), to provide a lower bound for their projections. However, the

difference between these extreme projections can yield large uncertainties (e.g., Thuiller et al. 2004).

Reducing these uncertainties requires taking dispersal processes into account more directly. While some studies have already included dispersal limitation when projecting species distribution under climate change scenarios (e.g., Dullinger et al. 2004, Iverson et al. 2004), none has done it for a large number of species at a fine spatial scale.

Using the MigClim cellular automaton (Engler and Guisan in press), which allows for implementation of plant dispersal constraints into projections of future species distribution, we here (i) project potential future distribution and assess the extinction risk over the 21st century for 287 plant species of the Western Swiss Alps while accounting for realistic dispersal limitations and, (ii) compare these results to those from unlimited and no-dispersal projections. More specifically, this study aims to address the following:

1) Are simulations implementing realistic dispersal limitations close to projections assuming no dispersal, unlimited dispersal or neither of these?
2) How much does the inclusion of long-distance dispersal events affect the projections?
3) What is the temporal pattern of plant species extinctions during the 21st century in our study area?
4) Does the predicted range reduction and extinction risk estimated with different dispersal simulations relate to species migration capacity and altitudinal optimum?

For each species, simulations were performed under four climate change and four dispersal scenarios: realistic dispersal distance with and without long-distance dispersal events, no dispersal, and unlimited dispersal.

Methods

Study area

The Diablerets study area (Fig.1a-b) is located in the Western Swiss Alps (6°50' to 7°15' E; 46°10' to 46°30' N) and covers 700 km^2. Elevation ranges from 375 m in Montreux to 3210 m at the top of the Diablerets massif. Mean annual temperature and yearly precipitation sum vary from 8°C and 1200 mm at 600 m to -5°C and 2600 mm at 3000 m (Bouët 1985). The soil parent material is mainly calcareous. Vegetation has been and is still influenced by human activity: pasture is commonly observed from the bottom of the Rhone valley to the sub-alpine and lower alpine zones. Meadows with varying levels of human management intensity are mainly found at lower elevations.

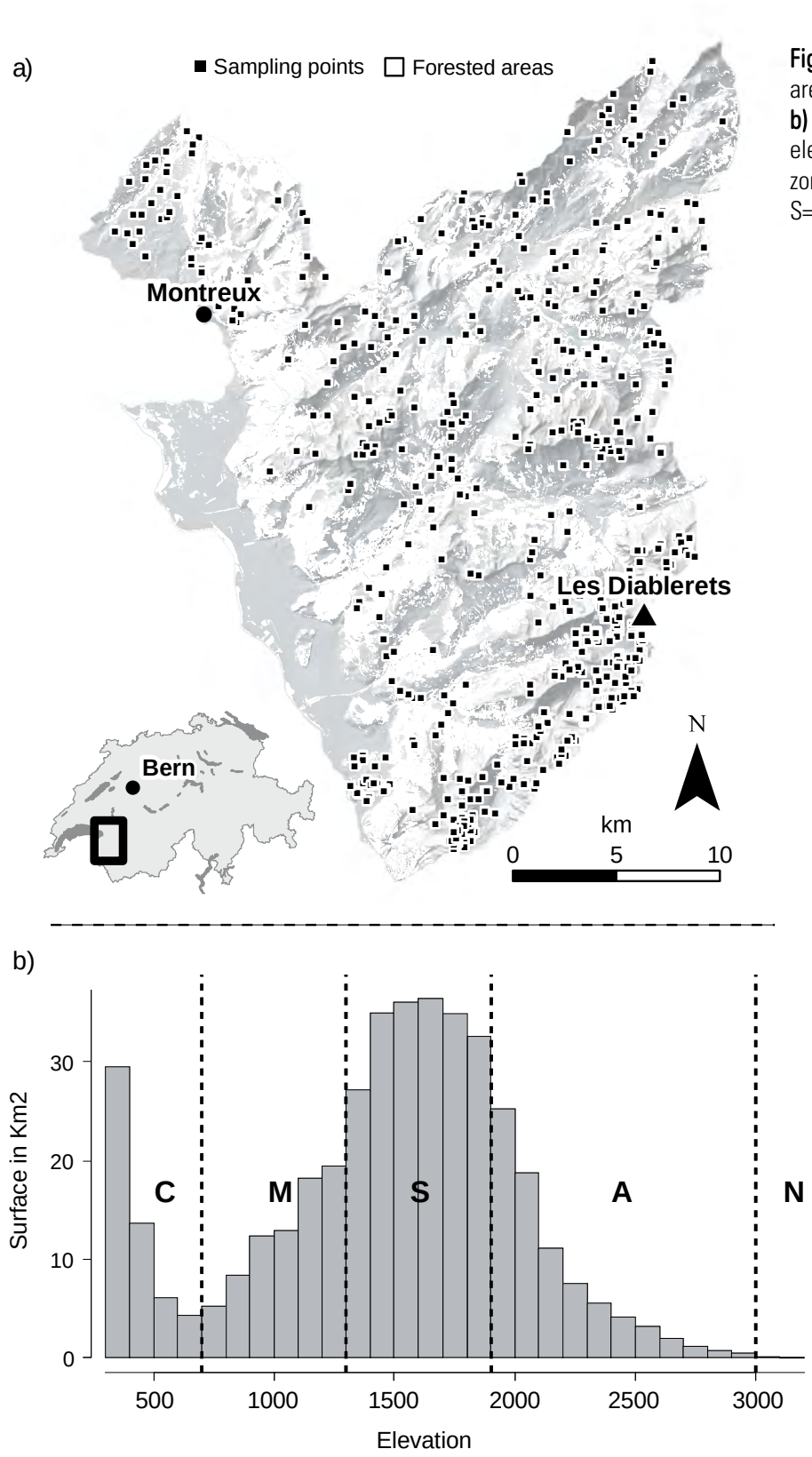


Figure 1. a) The Diablerets study area and its location in Switzerland. **b)** Frequency distribution of elevation with limits of vegetation zones: C=colline; M=montane; S=subalpine; A=alpine; N=nival.

Species dataset

During the summers of 2002-2004, 550 vegetation plots were sampled following a random-stratified sampling design restricted to non-woody vegetation (pastures and meadows below the timberline, grassland, rock, and scree above the line; Fig. 1a). Stratification included elevation, slope, aspect, and simplified classes of geology and land cover (Appendix S1, *supplementary material*). A minimum distance of 200 m was set between plots to minimize the effect of spatial auto-correlation when calibrating the models. Each plot had a 64 m^2 surface (8 × 8 meters) in which the presence of all species was recorded. Species with less than 20 presence points were removed to prevent loss of robustness in distribution models, leaving a total of 287 species.

Topographic and climatic predictors

Climate data were derived from long-term (1961-1990) monthly means for average temperature (°C) and sum of precipitation (mm) provided by the Swiss national meteorological station network at different altitudes. Methods for computation and description of the variables are summarized in Zimmermann and Kienast (1999). In this study, we employed three climatic and two topographic variables at a 25 m resolution (Table 1).

Table 1. Topographic and climatic variables used to model species distributions.

Variables	Units	Details	References
Temperature degree days	°C × day × $year^{-1}$	Sum of days multiplied by temperature > 0°C	Zimmermann and Kienast (1999)
Moisture index (average of monthly values June–August)	mm × day^{-1}	Monthly average of daily atmospheric H_2O balance	Zimmermann and Kienast (1999)
Global solar radiation (sum over the year)	kJ × m^{-2} × $year^{-1}$	Sum of monthly average of daily global solar radiation	Zimmermann et al. (2007)
Slope	degrees	Slope inclination	ArcGIS Spatial Analyst extension.
Topographic position	Unitless	Integration of topographic features (ridge, slope, toe slope) at various spatial scales	Zimmermann et al. (2007)

These variables are thought to best represent the physiological requirements of species, i.e., energy, water, and nutrients (Pearson et al. 2002, Körner 2003). Degree-days of the growing season (threshold of 0°C) were derived from interpolated monthly temperatures. The moisture index was calculated as the difference between precipitation and potential evapotranspiration, expressing the amount of soil water that is potentially available at a site. The sum of mean daily values for the months of June, July, and August were used, since these represent the warmest and driest months of the year. The potential global solar radiation was calculated over the year. Topographic position and slope were derived from a 25 meter resolution elevation model. Positive values of topographic position express relative ridges, tops, and exposed sites, while negative values indicate sinks, valleys, or toe slopes.

Climate change scenarios

We used four different climate projections for the 2001-2100 time period developed by the UK Hadley Center for Climate Prediction and Research. These were derived from a global circulation model (HadCM3; Carson 1999), and are based on four different socio-economic scenarios: A1Fi (hereafter referred to as "A1"), A2, B1, and B2 of the IPCC (Intergovernmental Panel on Climate Change; Nakicenovic and Swart 2000).

The climate change projections, in the form of 10' (~15km in Switzerland) grids, were used to calculate monthly mean temperature and precipitation anomalies for our study area between the standard period of 1961-1990 and 20 five-year periods from 2001 to 2100. These anomalies were then downscaled to 25 m resolution using bilinear interpolation and added to current monthly precipitations and temperatures. Thereafter, we calculated the derived environmental predictors for the study area under future climatic conditions. With an average increase of +7.6°C in our study area by 2100, the A1 climate change scenario is the most extreme. B1 is most modest (+ 3.9°C), while A2 (+6.2°C) and B2 (+4.4°C) are intermediate scenarios.

Species distribution models

For each species, Generalized Linear Models (GLM; McCullagh and Nelder 1989), Generalized Additive Models (GAM; Hastie and Tibshirani 1986), Gradient Boosting Machines (GBM; Friedman et al. 2000), and Random Forests (RF; Breiman 2001) were fitted using species presence-absence as the response variable and climatic and topographic variables as predictors (i.e., explanatory variables). GLMs and GAMs were calibrated using a binomial distribution and a logistic link function (i.e., logistic regression). A stepwise procedure in both directions was used for predictor selection, based on the Akaike information criterion (AIC; Akaike 1973). Up to second-order polynomials (linear and quadratic terms) were allowed for each predictor, with the linear term being forced into the model each time the quadratic term was retained.

The accuracy of each model was evaluated by 10-fold cross-validation (van Houwelingen and Le Cessie 1990) on the training data set. In the cross-validation procedure, the original prevalence of the species presences and absences in the data set was maintained in each fold. Comparisons of predicted (probabilistic scale) and observed (presence–absence) values were based on the area under the curve (AUC) of a receiver-operating characteristic plot (ROC; Fielding and Bell 1997).

Spatial projections

The modeling technique that produced, on average, the highest AUC values over the entire species dataset was selected to carry out spatial projections. Binary presence-absence distributions were derived from the probabilistic values of the niche-based models by maximizing the percentage of presences and absences that were correctly predicted under current climatic conditions when splitting probabilities (Pearce and Ferrier 2000).

Masks based on forests, lakes, urbanized areas, roads, and rivers were subsequently applied on the raw projections to avoid spurious projections at locations unsuitable for reasons other than climate. Finally, projections were further restricted to land cover classes (grasslands, rock, and scree) for which a species was observed at least once.

Dynamic simulation using the MigClim cellular automaton

MigClim (Engler and Guisan in press) is a cellular automaton that allows simulation of the dispersal of plant species in the landscape, while implementing a climate change scenario at the same time. The MigClim model allows distinguishing between regular dispersal (hereafter, referred to as short-distance dispersal), where most of the seeds (e.g., 99%) are distributed with a predictable pattern, and long-distance dispersal (LDD), which affects only a small number of seeds and relies on more stochastic means of dispersal. Further MigClim parameters that were used in the present study were generation time, resilience time to unfavorable conditions, and barriers to dispersal.

Because of the high number of species modeled in our study (i.e., 287), it is difficult to obtain very accurate data to calibrate the dispersal behavior of each species. In fact, such data does not exist for the large majority, if not all, of the species considered in our study. To overcome this lack of accurate data, we assigned each of our species to one of the seven dispersal types defined in Vittoz and Engler (2007) (Table 2, Appendix S2 and S3, *supplementary material*), based on morphological and seed dispersal traits found in Müller-Schneider (1986). This allowed to derive dispersal distances for each species that can be expected to lie within an accuracy of one order of magnitude, which is sufficient to achieve significant improvement over models that ignore dispersal (Engler and Guisan in press). Each category was also assigned a maximum dispersal distance for LDD events between 1 and 10 km (Table 2), which corresponds to values commonly found in previous studies (Nathan 2005 and references therein).

Table 2. Parameters used in the MigClim model for each dispersal category (1 to 7).

* Variations in LDD event frequencies reflect the corrections applied for pixel size and dispersal event frequency to maintain the 0.01 frequency assigned to 25 m pixels constant over all categories. ** Dispersal event frequency indicates the time between two successive dispersal events.

	Dispersal categories						
	1	2	3	4	5	6	7
Max disp. dist.	1 m	5 m	15 m	150 m	500 m	1500 m	5000 m
MaxLDD disp. dist.	1 km	1 km	1 km	1 km	5 km	5 km	10 km
LDD frequency*	0.002	0.002	0.0004	0.01	0.04	0.04	0.04
Disp. event freq.**	5 years		1 year				
Generation time	Parameter set individually for each species (see Appendix S3)						
Resilience time	Parameter set individually for each species (see Appendix S3)						
Barriers	Forests or nothing						
Filters	Urban areas, lakes, rivers, glaciers + grassland and/or rock and/or scree						
No of species	53	10	55	21	24	83	41
Pixel size	5 m			25 m	50 m		
No of pixels	30,000,000			4,800,000	1,200,000		

The decrease in colonization probability with distance from a source pixel was chosen to reflect a negative exponential seed dispersal kernel, a common shape (Willson 1993), although many others exist (e.g., Clark et al. 1999, Nathan and Muller-Landau 2000, Powell and Zimmermann 2004). This negative exponential kernel was used only to model regular seed dispersal, as LDD events were modeled as an additional, separate component. Details regarding the calibration of colonization probabilities are given in appendix S4 (*supplementary material*), and further explanation of the MigClim model can be found in Engler and Guisan (in press).

Species generation time (i.e., the time required before a newly colonized pixel can act as a seed source) and resilience to unfavorable environmental conditions (i.e., the time a pixel remains occupied even though its habitat has become unsuitable) were set individually for each species based on expert knowledge (Appendix S3). Landscape fragmentation was represented by forested areas that were used as "barriers" (i.e., pixels that impede dispersal across them, except for LDD events) as well as lakes, urbanized areas, roads and rivers that were considered "permanently unsuitable habitats" (i.e., pixels that are permanently unsuitable for the species, but allow dispersal across these areas).

For each species, four dispersal scenarios were run: no dispersal (= ND), realistic dispersal distance without LDD (= SDD for "short distance dispersal"), realistic dispersal distance with LDD (= LDD), and unlimited dispersal (= UD). All simulations were performed for the 100-year period of 2001-2100 and under each of the four IPCC climate change scenarios (A1, A2, B1, B2). The initial distribution of a species was defined as its potential habitat under current climatic conditions, i.e., 1961-1990 average. Changes in habitat suitability reflecting climate change were implemented every fifth year. Dispersal was simulated each year, except for categories 1 and 2 (Table 2), as these species have dispersal distances that are too short to be modeled on a yearly basis, even on a grid with 5 m resolution. Therefore, the dispersal distances of these categories were multiplied by five and dispersal was simulated every fifth year, which assumes a conservative best-case approximation, where spread rate = generation time × dispersal distance. Finally, because each simulation included some randomness, all simulations were repeated five times and results were averaged.

Species extinctions by 2100 and throughout the 21st century

We first calculated the percentage of species predicted to become extinct and quasi-extinct by 2100 for each combination of dispersal (i.e., ND, SDD, LDD and UD) and climate change scenario (i.e., A1, A2, B1, and B2). Extinction and quasi-extinctions were defined, respectively, as a decrease of 100% or 90% in potential distribution as compared to a species' initial area of occupancy. To examine the temporal pattern of extinctions, the percentage of species going extinct (100% threshold) was also computed for every five-year time period, from 2005 to 2100.

Relationship between species extinction risk, migration capacity, and altitudinal optimum.

The relationship between a decrease in the distribution of a species by 2100 (expressed as a percentage of its initial distribution) and two biological traits (i.e., migration capacity and altitudinal distribution optimum) was assessed using univariate linear regressions. Only species with a decreasing distribution were considered for these analyses.

Migration capacity was expressed through a migration rate index (Eq. 1):

$$Migration\ rate\ index = Log_{10}\left(\frac{MaxDispersal_{99\%\ of\ Seeds}}{Generation\ Time}\right) + 1 \qquad \text{(Eq. 1)}$$

where $MaxDispersal_{99\%\ of\ Seeds}$ is the maximum dispersal distance for 99% of the seeds (i.e., the short dispersal distance), and *Generation Time* is the number of years required for a species to start producing seeds.

An index was derived for each species to express altitudinal distribution optima. First, we assigned one of three possible frequency values (VA: 0 = absent, 1 = rare, 2 = commonly present) for each association of a species with an elevation belt (EB: 1=colline, 2=montane, 3=subalpine, 4=alpine, 5=nival), using data from the Atlas Flora Alpina (Aeschimann et al. 2005). The elevation optimum index was then calculated as (Eq. 2):

$$Elevation\ optimum\ index = \frac{\sum EB \times VA}{\sum VA} \qquad \text{(Eq. 2)}$$

This resulted in four categories: 1 < montane ≤ 2; 2 < subalpine ≤ 3; 3 < alpine ≤ 4; 4 < nival ≤ 5. Due to the low number of nival species, these were pooled with alpine species.

Results

Species distribution model performance.

AUC values showed that overall, GLM performed slightly better than GAM, GBM, and RF (Appendix S5, *supplementary material*). A Wilcoxon signed-rank test (samples grouped by species AUCs) also confirmed that AUC values were significantly higher for GLM than for GAM, GBM, and RF (p value < 0.001).

Projected distributions and species extinctions by 2100.

Among our pool of 287 species, the percentage of those going extinct in the study area (Fig. 2; 100% threshold) varied from 0.3% (UD under B1) up to 52.6% (ND under A1). Results

from simulations with realistic dispersal distances (i.e., SDD and LDD) yielded extinction rates ranging from ~1% under B1 to ~20% under A1. These extinction rates were always closer to those of UD than ND simulations.

Quasi-extinction rates (i.e., species with over a 90% decrease in distribution; Fig. 2), values reached up to 89% (ND under A1) and were never below 23% (UD under B1). SDD and LDD simulations produced fairly similar results, with quasi-extinction rates of approximately 70% under A1 (most extreme) and 30% under B1 (least extreme).

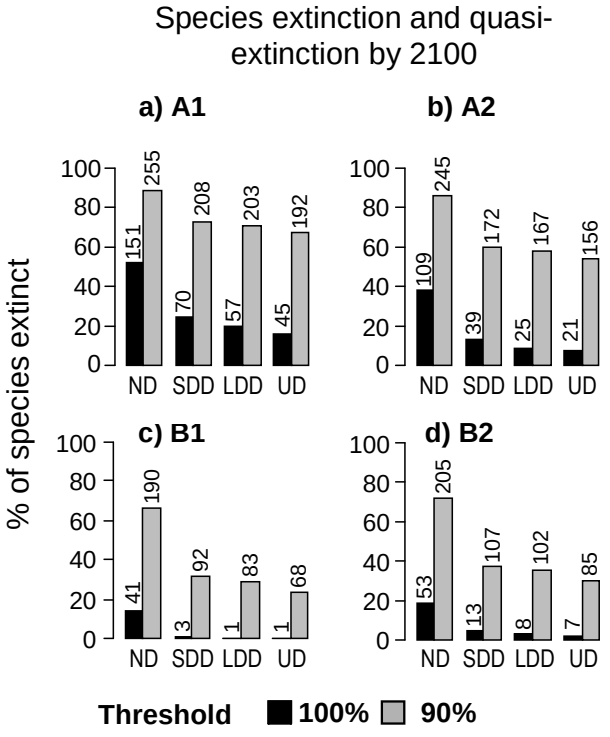


Figure 2. a-d) Percentages and absolute numbers (above histogram bars) of species predicted to become extinct (100% threshold) or quasi-extinct (90% threshold) by 2100 under the four climate changes (A1, A2, B1, B2) and dispersal scenarios (ND, SDD, LDD, UD). ND = no dispersal, SDD = short distance dispersal, LDD = long-distance dispersal, UD = unlimited dispersal.

Consideration of changes in potential spatial distribution rather than species extinctions or quasi-extinction reveals that at least 70% of the investigated species will likely decrease in distribution by the end of the 21st century (Fig. 3). For instance, even under B1 climate change scenario with unlimited-dispersal, a highly optimistic combination, 198 species out of 287 are forecasted to decrease in distribution. For over 50% of these species, the total area of occupancy decreases by at least 60% compared to their initial distribution (Fig. 3k, small arrow indicates the median value). Nevertheless, some species are predicted to significantly increase in spatial distribution (up to 500% of their initial area of occupancy). Comparison of the differences between ND, SDD, LDD, and UD simulations in terms of projected distribution surface for 2100 reveals that this difference is much larger for species with increases in potentially suitable habitat than those experiencing decreases (Fig 3e-g and 3i-n; compare mean values for "L" and "G"). For instance, under the A1 climate change scenario, the mean difference in surface between ND and SDD projections is 7% for species with decreasing suitable habitat and 132% for species with increasing suitable habitat (Fig 3e). Interestingly, the differences between ND and SDD/LDD are larger under milder climate change scenarios, while the differences between UD and SDD/LDD are larger under stronger climate change scenarios. Finally, similarly to extinction rates, the projections are generally closer to the UD than to the ND scenario, in terms of surface.

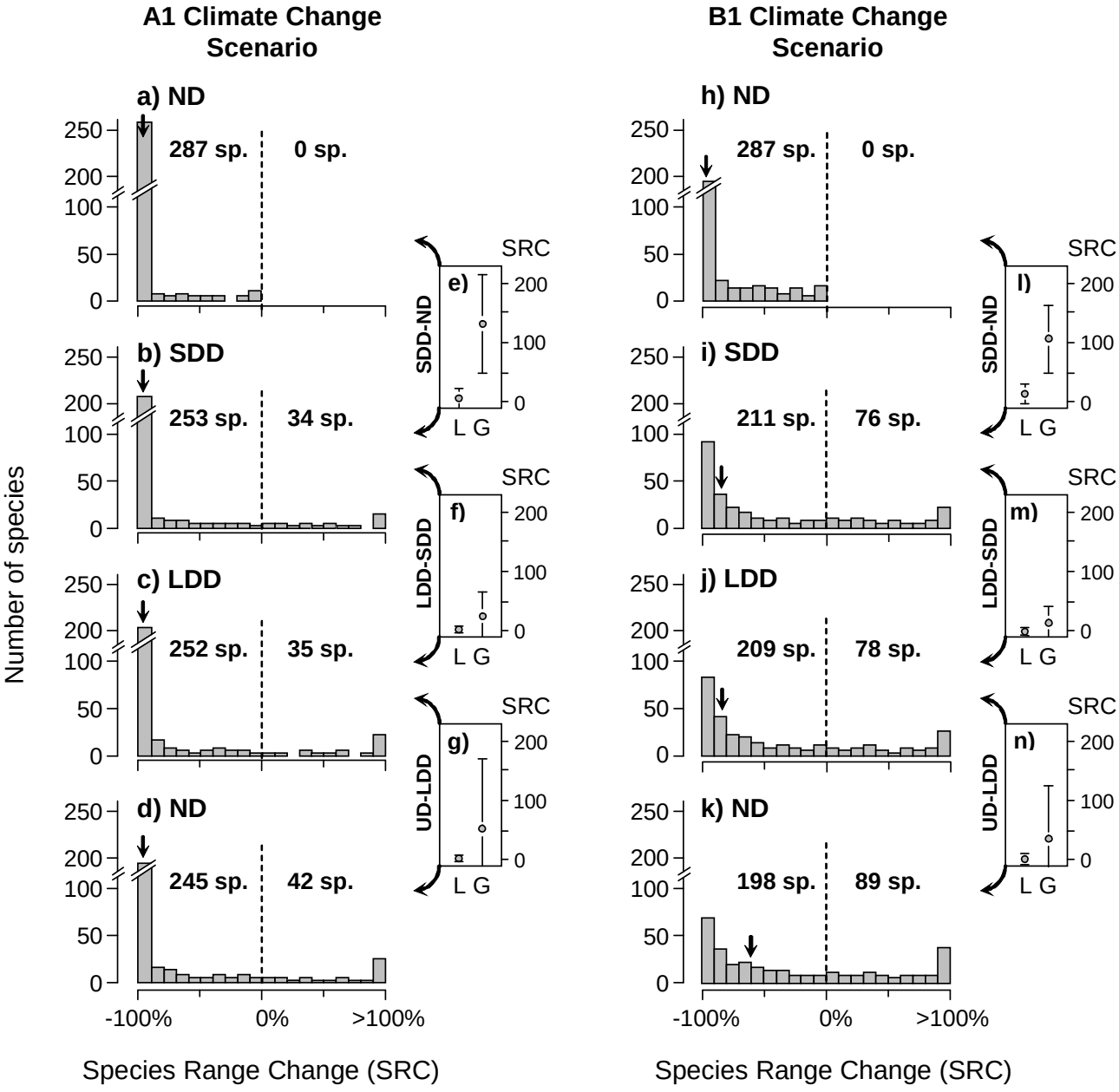


Figure 3. a-d) and **h-k)**: Frequency histograms of Species Range Change (SRC) for 2100 under A1 (most extreme; left panels) and B1 (least extreme; right panels) climate change scenarios. SRC is the change in distribution surface expressed as a percentage of a species' initially occupied area. Negative values indicate species that are projected to decrease in potential distribution and positive values species projected to increase in potential distribution. For instance, a value of -50% means that the predicted distribution by 2100 has, in terms of surface, decreased of 50% as compared to the species' initial distribution. The numbers given in each graph indicate the number of species loosing (SRC < 0; left number) or gaining (SRC > 0; right number) potential distribution. The small black arrow indicates the median value for all species.

e-g) and **l-n**): mean and standard deviation of the difference in SRC between ND and SDD (**e**, **l**), SDD and LDD (**f**, **m**), and LDD and UD (**g**, **n**) scenarios differentiated between potential species decreasing in distribution ("L" for "loss") and species increasing in distribution ("G" for "gain").

Projected species extinctions throughout the 21st century.

The graphs of cumulative species extinctions during the 21st century (Fig. 4) show that extinctions are expected to start occurring from about 2040 onwards under the A1 climate change scenario, and from 2080 - 2090 under the A2, B1, and B2 scenarios (except for ND scenarios where extinctions always start occurring already after 2040). In all scenarios, the large majority of extinctions are expected to occur between 2080 and 2100. The cumulative extinctions based on SDD and LDD simulations are always significantly lower than those obtained through the ND scenario (paired Wilcoxon signed-rank test, $p < 0.01$). Cumulative extinctions are also significantly different ($p < 0.05$) from each other under all climate change scenarios, except under B1 (the mildest), and are generally not significantly different from UD, except under A1 (the strongest).

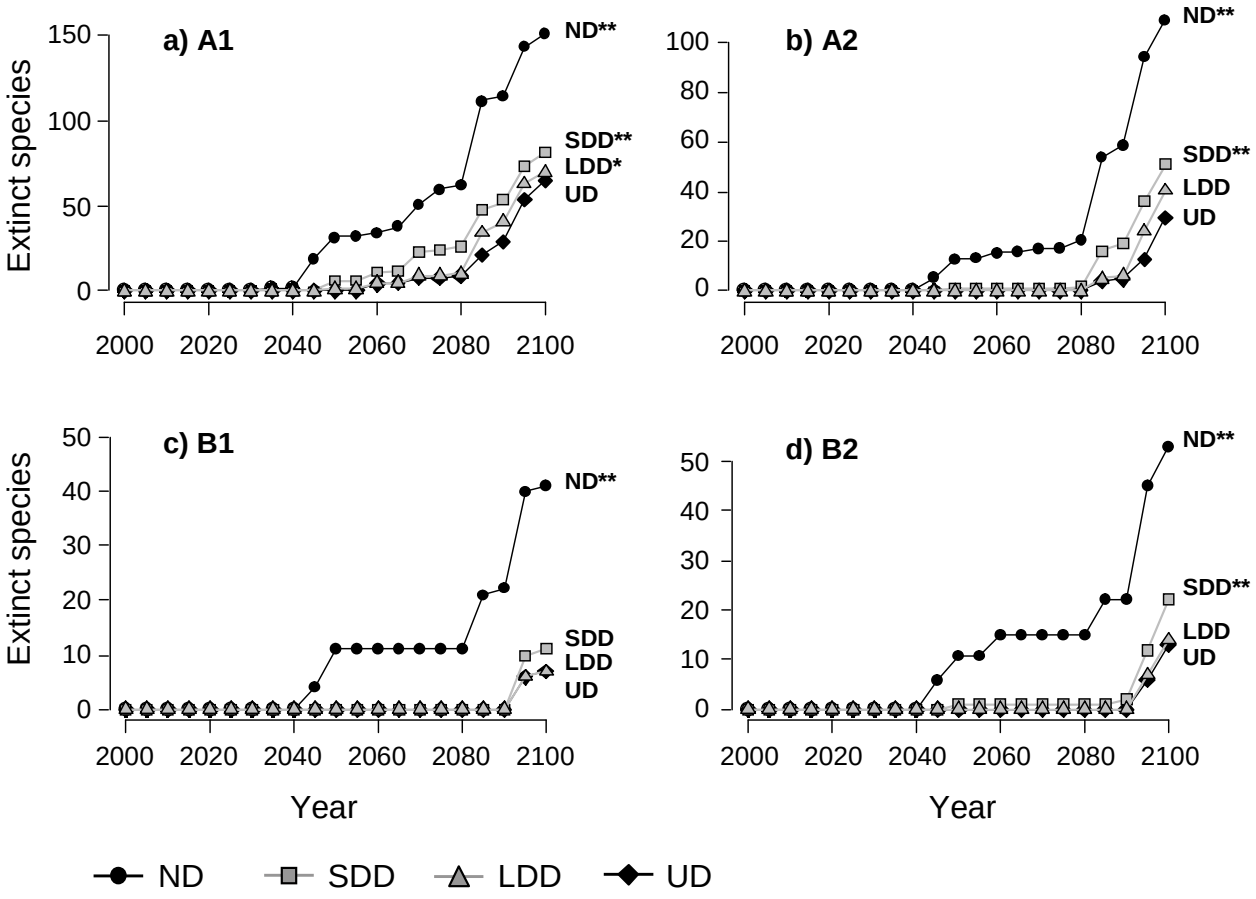


Figure 4. Cumulative number of species going extinct from 2000 to 2100 under the four climate change and four dispersal scenarios. ** = Series is significantly different from the series immediately below with $p < 0.01$, * = Series is significantly different from the series immediately below with $p<0.05$. Abbreviations as in Fig. 2.

Relationship between forecasted decrease in distribution and species traits.

Significant negative relationships exist between the migration index and decrease in potential distribution (Fig. 5a-b and Appendix S6, *supplementary material*), indicating that species with higher dispersal capacity appear less affected by climate change. These relationships between migration index and decrease in potential distribution are strongest for LDD and SDD simulations, but an important dispersion of observations along the linear trend was observed under all scenarios (adjusted R^2 = 0.14 at the most; SDD B1 scenario).

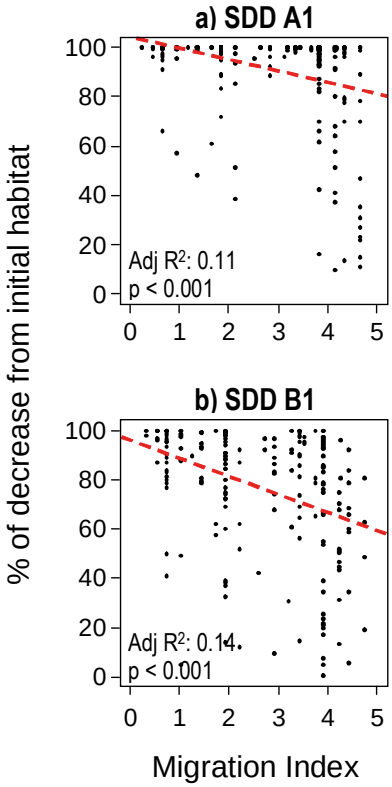


Figure 5. Relationship between migration rate index and decrease in projected distribution by 2100 shown for SDD simulations under A1 (a) and B1 (b) climate change scenarios. Explained variance (Adj. R^2) and significance of the regression slope (p-value) are given in each graph.

Under all climate change and dispersal scenarios, a significant positive relationship exists between the elevation optimum index and decrease in projected spatial distribution (Appendix S6). This relationship is stronger (higher adjusted R^2) for SDD, LDD, and UD than for ND simulations. This trend also appears in the distribution of extinctions within different elevation optima (Fig. 6): a higher proportion of sub-alpine than montane, and alpine than sub-alpine, species are predicted to go extinct.

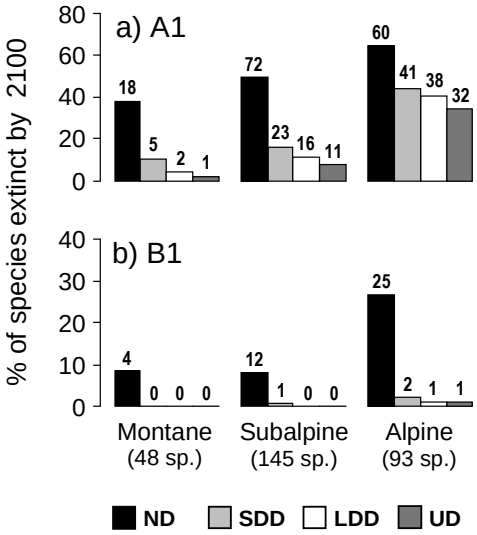


Figure 6. Percentage and absolute number (indicated above histogram bars) of species going extinct per altitudinal optimum for the A1 and B1 climate change scenarios. Abbreviations as in Fig. 2.

Discussion

In this study we projected the spatial distribution of 287 plants over the 21st century under four climate change scenarios in order to compare the results obtained from simulations with realistic dispersal limitations (SDD and LDD) to those assuming no or unlimited dispersal (ND and UD).

Are simulations accounting for dispersal limitations close to projections assuming no dispersal, unlimited dispersal, or neither?

Projections of SDD and LDD simulations were generally closer to UD than to ND, both in terms of species extinctions (Fig. 2) and surface of spatial distribution (Fig. 3). However, this does not necessarily indicate that UD is always a good approximation.

Considering the surface of future distribution, there is an important difference between those species for which suitable habitat is increasing as a result of climate change and those for which it is decreasing (Fig 3e-g and 3i-n; compare mean values for "L" and "G"). For the latter category, UD simulations generally provide fairly good approximations, with an average difference between SDD or LDD and UD of less than 5%. But for species whose suitable habitat increases, neither UD nor ND provided good approximations on average.

In terms of species extinctions, the difference between SDD or LDD and UD simulations can also remain important. For example, the extinction rate predicted by SDD simulations is ~24% under the A1 climate change scenario, but this value is only ~15% with the UD scenario (Fig. 2a). However, under milder climate change scenario (B1, B2), the difference between SDD or LDD and UD is small and the UD simplification can thus provide a relatively safe estimation of species extinctions. These results indicate that previous assessments of species extinctions performed under similar conditions (i.e., mountain area, local or regional scale, and fine pixel size), but assuming unlimited dispersal (e.g., Guisan and Theurillat 2000, Dirnböck et al. 2003), may have provided fairly correct – although slightly too low – estimates. For future studies in mountain environments at local or regional scales, indications that the unlimited dispersal simplification provides good estimations of species extinctions under moderate climate change scenario could save considerable time (i.e., accounting for dispersal limitations involves a significant amount of additional work).

Our results demonstrating that simulations accounting for dispersal limitations are closer overall to the unlimited-dispersal than to the no-dispersal scenario are most likely strongly linked to the fact that our study area is mountainous: for instance, Midgley et al. (2006) found very different, almost opposite, results when modeling *Proteaceae* of the Cape Floristic Region. The explanation for such differing results lies in the generally steeper climatic gradients of mountain regions, like our study area. Steeper gradients reduce the shift in suitable habitat when projected under climate change, and thus decrease the migration distance required to track those shifts. For this reason, we expect differences between SDD or LDD and UD scenarios to be much larger in non-mountainous areas than in our results. Furthermore, species in mountain habitats are more likely to experience a decrease in potentially suitable habitat over time rather than a shift or increase. In such cases, limitations in dispersal capacity have less impact on the projections, since there are not many new pixels to potentially colonize in the given study area.

How much do long-distance dispersal (LDD) events affect projected distributions?

The importance of LDD in plant migration remains a debated topic (Pearson 2006). On the one hand, several studies have shown that LDD can be important for determining plant dispersal rates (e.g., Higgins et al. 2003, Pearson and Dawson 2005, Soons and Ozinga 2005), and particularly to explain previous fast plant migrations (Clark 1998, Nathan and Muller-Landau 2000, Higgins et al. 2003, Powell and Zimmermann 2004). On the other hand, late-glacial refugia and molecular data indicate that migration rates might not have been as high as initially thought (see Pearson 2006 for a review), at least not in the Alps where mountains may have prevented expansions (Taberlet et al. 1998).

In our study, addition of LDD events to our simulations resulted in species extinction rates to decrease almost to the level of UD simulations (Figs. 2 and 4). While extinction rate differences between SDD and LDD simulations were sometimes significant, they were never very large. This can be explained by the study area configuration and the nature of the investigated species pool. First, the study area might be too small for LDD to have a high impact. Simulations made at a larger geographical scale are likely to exhibit greater differences between SDD and LDD scenarios. Second, most of our species undergo a decrease in potential spatial distribution. Thus, even if these species are given the possibility of longer dispersal distances, there is relatively modest suitable surface available. Note that for species that increase in distribution, including LDD can have a non-negligible impact. The mean increase in distribution surface due to LDD is between ~15% and ~25%, depending on the climate change scenario (Fig 3f and 3m).

What temporal patterns of extinctions are projected for the 21st century?

Projections of species distribution models are most often performed for environmental conditions of one or occasionally two future time periods, which are typically averages for several decades (e.g., Bakkenes et al. 2002, Dirnböck et al. 2003, Thomas et al. 2004, Thuiller et al. 2005). By projecting species extinctions for every five year period (Fig. 4) we are able to identify critical periods of expected extinctions within our study area over the 21st century. In our projections, species losses occur from 2050 onwards for the most pessimistic scenario (A1), and after 2080 for the three other climatic scenarios, except for ND simulations where species extinctions start at ~2040. This inertia in species extinctions has two explanations: either a species has the possibility to move upwards to higher elevation, thereby postponing extinction, or its initial distribution covers a wide altitudinal range, making it adapted to a broader range of environmental conditions, and therefore, more resilient to changes in these conditions. Clearly, the dates of 2040-2050 and 2080 can only be considered valid for our study area and species pool, and should not be extrapolated to other areas, although some inertia in species extinction is very likely to occur for other mountain areas too.

Temporally explicit projections, such as in this study, are important for conservation planning. For instance, the identification of areas where extinction is predicted to occur by 2040 and 2080 may be preserved from direct land use impacts in order to reduce future threats. It also highlights the fact that an absence of observed extinctions during the next 40 - 80 years should not be interpreted as an absence of the threat to species extinction in the longer term.

Does extinction risk and reduction in potential spatial distribution relate to species migration capacity and elevation optima?

We found significant relationships between the decrease in predicted distribution by 2100 and the species' migration capacity and altitudinal optimum (Fig. 5, 6 and Appendix S6). Overall, alpine plants and species with dispersal distances less than 20 meters appear to be particularly vulnerable to complete extinction by the end of the 21st century. This corroborates results from previous modeling studies (e.g., Guisan and Theurillat 2000, Dirnböck et al. 2003), and also from field-observations. For instance, Klanderud and Birks (2003) compared species compositions on Norwegian summits between 1930 and 1998, and reported that high elevation species had disappeared from lower elevation sites, with climate change as the most plausible explanation. In another study in the Italian Central Alps, Parolo and Rossi (2008) found that alpine species with larger seed weight and morphology less favorable for dispersal, migrated more slowly than other species over a 50 year period, suggesting that such species might be at greater risk under rapid climate change. Finally, Vittoz et al. (in press) also identified that there were more species with efficient dispersal modes among good colonizers of several alpine summits than among weak colonizers.

Projection limitations and other sources of uncertainty.

Since our habitat suitability projections were generated using GLMs, all limitations that are inherent to these models when used for climate change impact projections (e.g., truncated response curves, non-equilibrium of species with environment, see Guisan and Zimmermann 2000, Guisan and Thuiller 2005 for more details) also apply to our results. For instance, we need to be aware that our projections of extinction rates include some uncertainty, since the realized, and not the fundamental niche, of a given species was calibrated. Thus, if the realized and the fundamental niche of a target species differ clearly regarding the relevant changing climate factors, then our target species might actually be more tolerant than expected to predicted climate change over the shorter term. Over the longer term, we would – however – expect that this species is slowly out-competed by stronger competitors if the climate changes to low suitability of the realized niche. Future changes in land use (e.g., Vittoz et al. 2009) are a further source of uncertainty.

Furthermore, one should also keep in mind that the MigClim model and the data used herein for calibration remain a simplification compared to models that include more complex population dynamics and seed dispersal kernels (e.g., Dullinger et al. 2004, Powell and Zimmermann 2004, Pearson and Dawson 2005) or physiological response curves (Humphries et al. 1996). However, given that calibrating such complex models for hundreds of species is nearly impossible simply because accurate data on population dynamics and seed probability distribution functions are not easily available for the large majority of species, our approach seems to represent a good balance between accuracy and use of available data.

Another point worth mentioning is that, although accounting for dispersal limitations potentially reduces uncertainty in projections (i.e., as compared to the unlimited dispersal – no dispersal simplifications), the part of uncertainty related to the magnitude of climate change or the criteria chosen to declare a species threatened (e.g., 90% or 100% thresholds of surface loss) remains very significant (e.g., Fig. 2).

Finally, because dynamic simulations accounting for dispersal limitations depend on a large number of local parameters (e.g., habitat fragmentation, species migration rates) the

results provided by such models are necessarily site-specific, making their extrapolation to other study areas difficult.

Conclusion

In our study, simulations accounting for realistic dispersal limitations yielded extinction rates between ~1% (B1) and 25% (A1), and decreases in distribution for ~70% (B1) to 90% (A1) of the 287 species by the year 2100. Furthermore, we identified the following trends in our results:

- Restricting projections of future plant distribution through realistic dispersal limitations produced results that were generally closer to unlimited dispersal than to no dispersal.
- In terms of species extinctions, using the unlimited dispersal simplification provided a good approximation of the results obtained while accounting for dispersal limitations. This is especially true under less extreme climate change scenarios (B1, B2).
- In terms of projected distribution surface, the unlimited dispersal simplification provided fair approximations for species with decreasing distributions, but not for those whose distribution is increasing.
- Including long-distance dispersal (LDD) events into the simulations generally produced extinction rates that were very close to those of the unlimited dispersal scenario.
- The pattern of cumulative species extinction over time did not exhibit a linear increase. In simulations accounting for dispersal limitations, important numbers of extinctions are expected to start occurring for our study area during the 2080-2100 period.
- High elevation and/or slow dispersing species are the most at risk in the face of climate change.

Although our results are based on a fairly large number of species, they nevertheless remain very dependent on the species considered and on the study area. Therefore, generalizing our results to other areas should be done cautiously. While these results can probably be generalized to other mountainous study areas of similar size and shape, they cannot be extrapolated as such to larger and flatter landscapes. In fact, running such simulations at a continental scale or in larger and flatter study areas is likely to produce very different results. The influence of introducing migration limitations into the projections of species distribution under climate change in such environments remains to be tested.

Acknowledgements

This study was supported by the Centre de Conservation de la Nature (SFFN) of the Canton de Vaud, the Swiss National Science Foundation (SNF Grant No 3100A0-110000) and the European Commission (FP6 projects ECOCHANGE [GOCE-036866] and MACIS). We thank all

the people who helped with the field work, particularly Stéfanie Maire, Dario Martinoni, Séverine Widler, Chantal Peverelli, Valentine Hof, Grégoire Vuissoz, Lorenzo De Stefani, Roxane Milleret, Pierre Bermane Favrod-Coune, Daniel Cherix, Anny and Willy Berra. We also thank Hope C. Humphries, Jens-Christian Svenning and two anonymous reviewers for their comments on an early version of the manuscript, as well as Dorothea Pio for help with linguistic revision.

References

Aeschimann, D. et al. 2005. Flora Alpina. Haupt Verlag, Bern, Switzerland.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 in Second International Symposium on Information Theory. - Akademiai Kiado, Budapest.

Bakkenes, M. et al. 2002. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. - Glob. Chang. Biol. 8(4): 390-407.

Bouët, M. 1985. Climat et météorologie de la Suisse romande.

Breiman, L. 2001. Random Forests. - Machine Learning 45(1): 5-32.

Carson, D.J. 1999. Climate Modelling: achievements and prospects. - Quarterly Journal of the Royal Meteorological Society 125: 1-27.

Clark, J.S. 1998. Why trees migrate so fast: confronting theory with dispersal biology and the paleorecords. - Am. Nat. 152(2): 204-224.

Clark, J.S. et al. 1999. Seed Dispersal Near and Far: Patterns across temperate and tropical forests. - Ecology 80(5): 1475-1494.

Davis, M.B. and Shaw, R.G. 2001. Range Shifts and Adaptive Responses to Quaternary Climate Change. - Science 292: 673-679.

Dirnböck, T. et al. 2003. A regional impact assessment of climate and land-use change on alpine vegetation. - J. Biogeogr. 30(3): 401-417.

Dullinger, S. et al. 2004. Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. - J. Ecol. 92(2): 241-252.

Engler, R. and Guisan, A. 2009. MigClim: Predicting plant distribution and dispersal in a changing climate. - Divers. Distrib. 15: 590-601.

Eriksson, O. 2000. Functional roles of remnant plant populations in communities and ecosystems. - Glob. Ecol. Biogeogr. 9: 443-449.

Etterson, J.R. and Shaw, R.G. 2001. Constraint to Adaptive Evolution in Response to Global Warming. - Science 294: 151-154.

Fielding, A.H. and Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence-absence models. - Environ. Conserv. 24(1): 38-49.

Folland, C.K. et al. 2001. Global temperature change and its uncertainties since 1861. - Geophysical Research Letters 28(13): 2621-2624.

Friedman, J. et al. 2000. Additive logistic regression: a statistical view of boosting. - Ann. Stat. 28(2): 337-407.

Guisan, A. and Theurillat, J.-P. 2000. Assessing alpine plant vulnerability to climate change: A modeling perspective. - Integrated Assessment 1: 307-320.

Guisan, A. and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. - Ecol. Model. 135(2-3): 147-186.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. - Ecol. Lett. 8: 993-1009.

Hastie, T. and Tibshirani, R. 1986. Generalized Additive Models. - Statistical Science 1(3): 297-318.

Higgins, S.I. et al. 2003. Are long-distance dispersal events in plants usually caused by nonstandard means of dispersal ? - Ecology 84(8): 1945-1956.

Humphries, H.C. et al. 1996. An individual-based model of alpine plant distributions. - Ecol. Model. 84: 99-126.

IPCC (ed.). 2007. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment. Report of the Intergovernmental Panel on Climate Change. - Cambridge University Press.

Iverson, L.R. et al. 2004. How fast and far might tree species migrate in the eastern United States due to climate change? - Glob. Ecol. Biogeogr. 13(3): 209-219.

Jump, A.S. and Peñuelas, J. 2005. Running to stand still: adaptation and the response of plants to rapid climate change. - Ecol. Lett. 8: 1010-1020.

Klanderud, K. and Birks, H.J.B. 2003. Recent increases in species richness and shifts in altitudinal distributions of Norwegian mountain plants. - The Holocene 13: 1-6.

Körner, C. 2003. Alpine plant life. 2nd Ed. - Springer.

Loarie, S.R. et al. 2008. Climate Change and the Future of California's Endemic Flora. - PLoS One 3(6): e2502.

Malcolm, J.R. et al. 2002. Estimated migration rates under scenarios of global climate change. - J. Biogeogr. 29: 835-849.

McCullagh, P. and Nelder, J.A. 1989. Generalized Linear Models, 2nd edn. - Chapman & Hall, London, UK.

Midgley, G.F. et al. 2006. Migration rate limitations on climate change-induced range shifts in Cape Proteaceae. - Divers. Distrib. 12: 555-562.

Müller-Schneider, P. 1986. Verbreitungsbiologie der Blütenpflanzen Graubündens. - Veröff. Geobot. Inst. ETH Stiftung Rübel Zürich 85: 1-263.

Nakicenovic, N. and Swart, R. (eds.). 2000. Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change. - Cambridge University Press.

Nathan, R. 2005. Long-distance dispersal research: building a network of yellow brick roads. - Divers. Distrib. 11: 125–130.

Nathan, R. and Muller-Landau, H.C. 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. - Trends Ecol. Evol. 15(7): 278-285.

Parolo, G. and Rossi, G. 2008. Upward migration of vascular plants following a climate warming trend in the Alps. - Basic Appl. Ecol. 9: 100-107.

Pearce, J. and Ferrier, S. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. - Ecol. Model. 128(2-3): 127-147.

Pearson, R.G. 2006. Climate change and the migration capacity of species. - Trends in Ecology & Evolution 21(3): 111-113.

Pearson, R.G. and Dawson, T.P. 2005. Long-distance plant dispersal and habitat fragmentation: identifying conservation targetsfor spatial landscape planning under climate change. - Biol. Conserv. 123: 389-401.

Pearson, R.G. et al. 2002. SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. - Ecol. Model. 154(3): 289-300.

Pitelka, L.F. et al. 1997. Plant migration and climate change. - Am. Sci. 85(5): 464-473.

Powell, J.A. and Zimmermann, N.E. 2004. Multi-scale analysis of seed dispersal and the resolution of Reid's Paradox. - Ecology 85(2): 490-506.

Soons, M.B. and Ozinga, W.A. 2005. How important is long-distance seed dispersal for the regional survival of plant species ? - Divers. Distrib. 11: 165-172.

Taberlet, P. et al. 1998. Comparative phylogeography and postglacial colonization routes in Europe. - Mol. Ecol. 7: 453-464.

Thomas, C.D. et al. 2004. Extinction risk from climate change. - Nature 427(6970): 145-147.

Thuiller, W. et al. 2004. Uncertainty in predictions of extinction risk. - Nature 430: 34.

Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. - Proc. Nat. Acad. Sci. 102(23): 8245-8250.

van Houwelingen, J.C. and Le Cessie, S. 1990. Predictive value of statistical models. - Stat. Med. 9: 1303-1325.

Vittoz, P. and Engler, R. 2007. Seed dispersal distances: a typological system for data analyses and models. - Bot. Helv. 117: 109-124.

Vittoz, P. et al. 2009. Low impact of climate change on subalpine grasslands in the Swiss Northern Alps? - Glob. Chang. Biol. 15: 209-220.

Vittoz, P. et al. in press. Plant traits discriminate good from weak colonisers on high elevation summits. - Basic Appl. Ecol.

Willson, M.F. 1993. Dispersal Mode, Seed Shadows, and Colonization Patterns. - Vegetatio 108: 261-280.

Zimmermann, N.E. and Kienast, F. 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. - J. Veg. Sci. 10(4): 469-482.

Supplementary Material / Supporting Information

Appendix S1

Variables and classes used in the random stratified sampling design.

Variable	Number of classes	Classes
Elevation	4	Colline (<700m)
		Montane (700-1300m)
		Subalpine (1300-1900m)
		Alpine+nival (>1900m)
Aspect	2	North slopes (0-90°+270-360°)
		South slopes (90-270°)
Slope	3	0-10°
		10-45°
		45-90°
Geology	4	Calcareous permeable
		Calcareous unpermeable
		Other rock types permeable
		Other rock types unpermeable
Land cover	3	Exploited (i.e. pasture)
		Unexploited meadows
		Rocks and screes

Appendix S2

Dispersal categories as defined by Vittoz and Engler (2007). The given Dispersal distance corresponds to the maximum dispersal distance of 99% of the seeds, which thus excludes long-distance dispersal events. For details on how these categories were obtained see Vittoz and Engler (2007).

Type	Dispersal distance [m] of 99% of seeds.	Corresponding dispersal modes
1	1	Blastochory (autochory). Boleochory (anemochory) for species < 30 cm. Ombrochory (hydrochory).
2	5	Ballochory (autochory). Cystometeorochory (anemochory). Chamaechory (anemochory) for fruits in grassland. Boleochory (anemochory) for species > 30 cm.
3	15	Pterometeorochory (anemochory) for herbs. Myrmecochory (zoochory). Cystometeorochory (anemochory) ferns, Orchidaceae, Pyrolaceae, Orobanchaceae in forest. Trichometeorochory (anemochory) in forest or little efficient plumes. Epizoochory (zoochory) for small mammals.
4	150	Chamaechory (anemochory) for seeds on snow or dry inflorescence. Pterometeorochory (anemochory) for trees. Dyszoochory (zoochory) for seeds not stocked and dispersed by small animals.
5	500	Trichometeorochory (anemochory) in openland with efficient plumes. Cystometeorochory (anemochory) ferns, Orchidaceae, Pyrolaceae, Orobanchaceae in openland.
6	1500	Dyszoochory (zoochory) for seeds stocked by large animals. Endozoochory (zoochory) for seeds eaten by birds and large vertebrates. Epizoochory (zoochory) by large mammals.
7	5000	Agochory (anthropochory).

References

Vittoz, P. and Engler, R. 2007. Seed dispersal distances: a typological system for data analyses and models. - *Botanica Helvetica,* **117**, 109-124.

Appendix S3

Dispersal category, generation time and resilience time assigned to each of the 287 species used in the study. Nomenclature follows Aeschimann and Heitz (1996).

Species Name	Dispersal Category	Generation Time [Years]	Resilience Time [years]
Acer pseudoplatanus	4	30	50
Achillea atrata	1	3	5
Achillea millefolium	6	2	5
Acinos alpinus	6	2	5
Aconitum napellus aggr.	1	2	5
Adenostyles glabra	5	2	5
Agrostis alpina	3	2	5
Agrostis capillaris	6	1	5
Agrostis rupestris	3	2	5
Agrostis schraderiana	3	2	5
Agrostis stolonifera	6	1	5
Ajuga reptans	3	1	5
Alchemilla conjuncta aggr.	6	2	5
Alchemilla coriacea aggr.	6	1	5
Alchemilla decumbens aggr.	6	2	5
Alchemilla glabra aggr.	6	1	5
Alchemilla xanthochlora aggr.	6	2	5
Androsace chamaejasme	4	2	5
Anemone narcissiflora	3	2	5
Anthoxanthum odoratum aggr.	7	2	5
Anthriscus sylvestris	6	2	5
Anthyllis vulneraria s.l.	6	1	5
Aposeris foetida	3	2	5
Arabis alpina s.str.	1	2	5
Arabis caerulea	1	2	5
Arabis hirsuta	1	1	5
Arnica montana	3	2	5
Arrhenatherum elatius	7	2	5
Asplenium viride	5	2	5
Aster bellidiastrum	3	2	5
Astrantia major	6	2	5
Athamanta cretensis	1	2	5
Bartsia alpina	1	2	5
Bellis perennis	6	2	5
Biscutella laevigata	3	1	5
Botrychium lunaria	5	2	5
Brachypodium pinnatum	3	1	5
Briza media	3	2	5
Bromus erectus s.str.	4	2	5
Bromus hordeaceus	7	1	0
Calamagrostis varia	5	2	5
Campanula barbata	1	2	5
Campanula cochleariifolia	1	2	5
Campanula rhomboidalis	1	1	5
Knautia arvensis	7	2	5
Knautia dipsacifolia s.str.	3	2	5
Laserpitium latifolium	3	3	5
Lathyrus pratensis	6	2	5
Leontodon autumnalis	3	2	5
Leontodon hispidus s.l.	3	2	5
Leucanthemum vulgare aggr.	6	2	5
Ligusticum mutellina	6	2	5
Linaria alpina s.str.	1	2	5
Linum alpinum	1	2	5
Linum catharticum	6	1	0
Lolium perenne	6	1	5
Lotus alpinus	6	2	5
Lotus corniculatus	6	1	5
Luzula alpinopilosa	6	2	5
Luzula campestris	6	2	5
Luzula multiflora	6	2	5
Luzula sylvatica	3	2	5
Medicago lupulina	6	1	0
Myosotis alpestris	4	2	5
Myosotis arvensis	7	1	0
Myosotis decumbens	6	1	5
Nardus stricta	7	2	5
Nigritella rhellicani	5	2	5
Oxytropis jacquinii	2	2	5
Parnassia palustris	2	2	5
Pedicularis foliosa	7	2	5
Pedicularis verticillata	1	3	5
Petasites paradoxus	5	2	5
Phleum alpinum aggr.	3	2	5
Phleum hirsutum	3	2	5
Phleum pratense	6	1	5
Phleum rhaeticum	3	2	5
Phyteuma orbiculare	6	2	5
Phyteuma spicatum	1	2	5
Picea abies	6	40	50
Pimpinella major	7	2	5
Pimpinella saxifraga aggr.	7	2	5
Plantago alpina	6	2	5
Plantago atrata s.str.	6	2	5
Plantago lanceolata	7	1	5
Plantago major s.str.	6	1	5
Plantago media	6	2	5
Poa alpina	6	2	5

Campanula rotundifolia	6	1	5
Campanula scheuchzeri	1	2	5
Cardamine pratensis	2	2	5
Carduus defloratus s.str.	5	2	5
Carex caryophyllea	6	2	5
Carex ferruginea	6	3	5
Carex flacca	4	2	5
Carex montana	3	2	5
Carex nigra	2	2	5
Carex ornithopoda	3	2	5
Carex pallescens	3	2	5
Carex panicea	2	2	5
Carex sempervirens	6	3	5
Carex sylvatica	6	2	5
Carlina acaulis subsp. caulescens	4	2	5
Carum carvi	6	1	0
Centaurea jacea s.str.	6	1	5
Centaurea montana	3	2	5
Centaurea scabiosa s.l.	4	1	5
Cerastium arvense s.l.	1	1	5
Cerastium fontanum subsp. vulgare	7	1	5
Cerastium latifolium	1	3	5
Chaerophyllum aureum	1	2	5
Chaerophyllum hirsutum aggr.	1	2	5
Cirsium acaule	5	2	5
Cirsium eriophorum s.str.	5	2	0
Cirsium palustre	5	2	0
Cirsium spinosissimum	5	2	5
Clinopodium vulgare	4	2	5
Coeloglossum viride	5	2	5
Colchicum autumnale	7	2	5
Crepis aurea	4	2	5
Crepis biennis	7	2	5
Crepis pyrenaica	7	2	5
Crocus albiflorus	7	2	5
Cruciata laevipes	7	2	5
Cynosurus cristatus	6	2	5
Cystopteris fragilis	5	2	5
Dactylis glomerata	7	1	5
Dactylorhiza maculata	5	2	5
Daphne mezereum	6	5	10
Daucus carota	6	2	0
Deschampsia cespitosa	7	2	5
Doronicum grandiflorum	5	3	5
Draba aizoides	1	2	5
Dryas octopetala	3	2	5
Elyna myosuroides	6	2	5
Epilobium anagallidifolium	5	1	5
Erigeron uniflorus	5	2	5
Euphorbia cyparissias	7	1	5
Euphrasia hirtella	1	1	0
Euphrasia minima	1	1	0
Euphrasia rostkoviana s.str.	1	1	0

Poa cenisia	6	2	5
Poa minor	6	2	5
Poa pratensis	7	1	5
Poa supina	7	1	5
Poa trivialis s.str.	7	1	5
Polygala alpestris	3	2	5
Polygala chamaebuxus	3	3	5
Polygala vulgaris s.str.	3	2	5
Polygonum bistorta	7	2	5
Polygonum viviparum	1	2	5
Polystichum lonchitis	3	2	5
Potentilla aurea	6	2	5
Potentilla crantzii	1	2	5
Potentilla erecta	4	1	5
Potentilla sterilis	3	2	5
Primula auricula	1	2	5
Primula elatior s.str.	6	2	5
Primula veris s.l.	1	2	
Pritzelago alpina s.str.	1	2	5
Prunella grandiflora	6	2	5
Prunella vulgaris	6	2	5
Pulsatilla alpina s.str.	3	2	5
Ranunculus aconitifolius	2	2	5
Ranunculus acris s.l.	7	1	5
Ranunculus alpestris	6	2	5
Ranunculus bulbosus	4	2	5
Ranunculus montanus aggr.	3	2	5
Ranunculus nemorosus aggr.	6	2	5
Ranunculus repens	3	1	5
Rhinanthus alectorolophus	7	1	0
Rhinanthus minor	1	1	0
Rhododendron ferrugineum	1	5	10
Rosa pendulina	6	3	10
Rubus idaeus	6	2	5
Rumex acetosa	6	1	5
Rumex alpestris	7	2	5
Rumex alpinus	6	2	5
Rumex crispus	4	2	5
Sagina saginoides	6	2	5
Salix reticulata	5	3	5
Salix retusa	5	3	5
Salvia pratensis	4	2	5
Sanguisorba minor s.str.	2	2	5
Saxifraga aizoides	1	3	5
Saxifraga androsacea	1	2	5
Saxifraga moschata s.l.	4	3	5
Saxifraga oppositifolia	1	3	5
Saxifraga paniculata	4	3	5
Scabiosa lucida	3	2	5
Sedum atratum	1	1	0
Selaginella selaginoides	5	2	5
Senecio doronicum	3	2	5
Sesleria caerulea	6	2	5

Euphrasia salisburgensis	1	1	0
Festuca arundinacea s.l.	3	1	5
Festuca ovina aggr.	3	2	5
Festuca pratensis s.l.	3	1	5
Festuca quadriflora	4	2	5
Festuca rubra aggr.	7	1	5
Festuca violacea aggr.	3	2	5
Fragaria vesca	6	2	5
Fraxinus excelsior	4	20	50
Galeopsis tetrahit	7	1	0
Galium album	6	2	5
Galium anisophyllon	6	1	5
Galium megalospermum	1	3	5
Galium pumilum	6	1	5
Gentiana acaulis	1	2	5
Gentiana bavarica	1	3	5
Gentiana campestris s.str.	6	2	0
Gentiana clusii	1	2	5
Gentiana lutea	4	2	5
Gentiana purpurea	1	2	5
Gentiana verna	1	2	5
Geranium sylvaticum	6	2	5
Geum montanum	3	3	5
Geum rivale	6	2	5
Geum urbanum	6	1	5
Glechoma hederacea s.str.	3	1	5
Globularia cordifolia	2	2	5
Globularia nudicaulis	2	2	5
Gymnadenia conopsea	5	3	5
Gypsophila repens	1	2	5
Hedysarum hedysaroides	3	2	5
Helianthemum nummularium s.l.	6	2	5
Helictotrichon pubescens	3	2	5
Helictotrichon versicolor	4	2	5
Heracleum sphondylium s.l.	7	3	5
Hieracium bifidum aggr.	3	2	5
Hieracium lactucella	3	2	5
Hieracium murorum aggr.	3	2	5
Hieracium villosum aggr.	3	2	5
Hippocrepis comosa	6	2	5
Holcus lanatus	6	2	5
Homogyne alpina	3	2	5
Hypericum maculatum s.str.	1	2	5
Hypericum perforatum s.str.	4	1	5
Hypochaeris radicata	7	2	5
Juncus effusus	1	2	5
Juniperus communis subsp. nana	6	5	20
Silene acaulis	1	5	5
Silene dioica	6	1	0
Silene vulgaris s.str.	7	1	5
Soldanella alpina	1	2	5
Solidago virgaurea s.l.	3	2	5
Stachys officinalis s.str.	1	2	5
Stellaria graminea	1	1	5
Taraxacum alpinum aggr.	5	2	5
Taraxacum officinale aggr.	7	1	5
Thesium alpinum	6	2	5
Thlaspi repens	1	2	5
Thymus praecox subsp. polytrichus	6	2	5
Thymus pulegioides s.str.	6	2	5
Tofieldia calyculata	1	2	5
Tragopogon pratensis subsp. orientalis	7	2	5
Traunsteinera globosa	5	3	5
Trifolium badium	6	2	5
Trifolium medium	6	2	5
Trifolium montanum	6	2	5
Trifolium pratense s.str.	6	1	5
Trifolium repens s.str.	7	1	5
Trifolium thalii	6	2	5
Trisetum flavescens	3	2	5
Trollius europaeus	7	2	5
Tussilago farfara	5	2	5
Urtica dioica	7	1	5
Vaccinium gaultherioides	6	5	5
Vaccinium myrtillus	7	3	5
Vaccinium vitis-idaea	7	3	5
Valeriana montana	3	2	5
Valeriana officinalis aggr.	3	2	5
Valeriana tripteris	3	2	5
Veratrum album subsp. lobelianum	2	3	5
Veronica alpina	1	2	5
Veronica aphylla	1	2	5
Veronica arvensis	7	1	0
Veronica chamaedrys	1	2	5
Veronica officinalis	6	2	5
Veronica persica	3	1	0
Veronica serpyllifolia s.l.	6	1	0
Vicia cracca s.str.	7	1	5
Vicia sativa s.l.	7	1	0
Vicia sepium	6	1	5
Viola biflora	4	2	5
Viola calcarata	3	2	5
Viola hirta	3	2	5

References

Aeschimann, D. and Heitz, C. 1996. Index synonymique de la flore de Suisse et territoires limitrophes. Documenta Floristicae Helvetiae.

Appendix S4

Calibration of P_{Disp}, the colonization probability of a pixel given its distance from a source pixel. This methodology is the same than the one developed in Engler and Guisan (in press).

The first step in calibration P_{Disp} is to define the dispersal kernel used to model regular seed dispersal (i.e., not including LDD). Here this kernel was based on the following negative exponential seed dispersal probability distribution function (Eq. S1; Ward *et al*. 2004; Scheller *et al*. 2007).

$$P_{seed}(x) = e^{(x-pixelsize)\cdot\left(\frac{\ln(1-k)}{DispDist}\right)} - e^{x\cdot\left(\frac{\ln(1-k)}{DispDist}\right)} \qquad \text{(Eq. S1)}$$

which can be simplified into the more conventional simple negative exponential form (Eq. S2):

$$P_{seed}(x) = \left(\left(1-k\right)^{-\frac{pixelsize}{DispDist}} - 1\right)\cdot e^{x\cdot\left(\frac{\ln(1-k)}{DispDist}\right)} \qquad \text{(Eq. S2)}$$

where P_{seed} is the probability of a seed reaching distance $x \geq pixelsize$, *pixelsize* is the one-dimensional size of a pixel, *DispDist* is the dispersal distance reached by the proportion k of the seeds. Here we set k = 0.99, *DispDist* thus represents the regular dispersal distance for seeds, i.e., LDD events excluded (these are modelled as a separate process).

Since the surface composed of pixels located at distance j from a source cell increases with distance from that source cell, the probability of a pixel to receive a seed is computed as (Eq. S3):

$$P_{seed}(pixel_j) = {P_{seed}(x)}\Big/{Surface_j} \qquad \text{(eq. S3)}$$

where $Surface_j$ is the number of pixels covered by all pixels belonging to a same distance class. Assuming that the distribution of successful seeds (i.e., seeds leading a pixel to become colonized) is proportional to the overall distribution of seeds (P_{seed}), P_{Disp} is computed as (eq. S3):

$$P_{Disp}(pixel_j) = 1-(1-P_{seed}(pixel_j))^{Successful\,Seeds} \qquad \text{(eq. S4)}$$

where P_{Disp} is the probability of colonisation for a target pixel with distance j from a source pixel and *Successful Seeds* the number of successful seeds produced by a fully mature source pixel.

Since the value of *Successful Seeds* was unknown for our virtual species (as is the case for most species), it was set so that P_{Disp} = 0.99 for a pixel immediately adjacent to a source pixel at 25 meters resolution. Using this conservative (i.e., optimistic) calibration method led to

average spread rates, for fully mature pixels, between 45% and 85% of *DispDist*, depending on the number of source pixels in the neighbourhood (tests run on homogenous landscapes). When running simulations at smaller pixels sizes (i.e., 12.5 m and 5 m), the values of P_{Disp} were recomputed in order to ensure that the production of *Successful Seeds* per surface unit remained constant. In other words, the number of *Successful Seeds* was always proportional to pixel size, 5 m and 12.5 m pixel producing respectively 25 and 4 times less successful seeds than 25 m pixels. This ensured that the species had always the same seed production per surface unit and thus that their spread rate was independent of the cell size at which simulations were run.

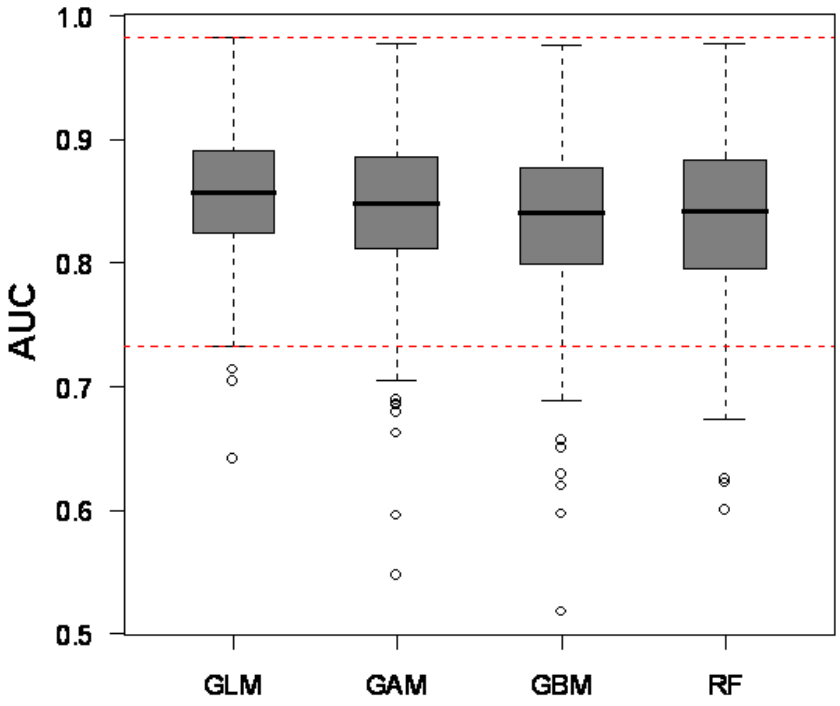
For more details on the MIGCLIM model and its parameters see Engler and Guisan (2009).

References

Engler R. and Guisan A., 2009. MigClim: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, **15**, 590-601.

Appendix S5

Boxplots of 10-fold cross-validated AUC values for GLM, GAM, GBM and RF models. Wilcoxon matched pair signed-rank test showed that AUC values are significantly higher for GLM (*P* values < 0.001) when compared to GAM, GBM and RF. The line across the box indicates the median, box boundaries show the interquartile range and whiskers extend up to 1.5 times the interquartile range. Horizontal dashed lines indicate whiskers position of GLM model.



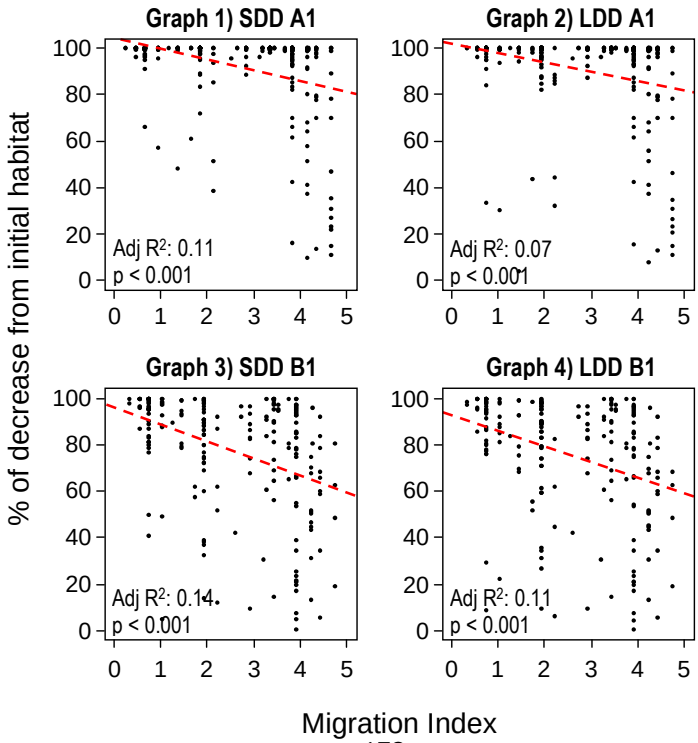
Appendix S6

Adjusted explained deviance (Adj. R^2) and regression slope p-value for regressions of decrease in predicted distribution by 2100 vs. **a)** Migration index and **b)** Elevation index. Only species with decreasing predicted distribution were taken into account (the number of species considered is given in the last column of each table). Examples of regression graphs for SDD and LDD simulations under A1 (most extreme) and B1 (least extreme) climate change scenario are given under each table. The yellow coloured rows indicate that these regressions are illustrated by a graphic (below) and the superscript indicates the precise graph the regression refers to.

ND = no dispersal, SDD = short distance dispersal, LDD = long-distance dispersal, UD = unlimited dispersal.

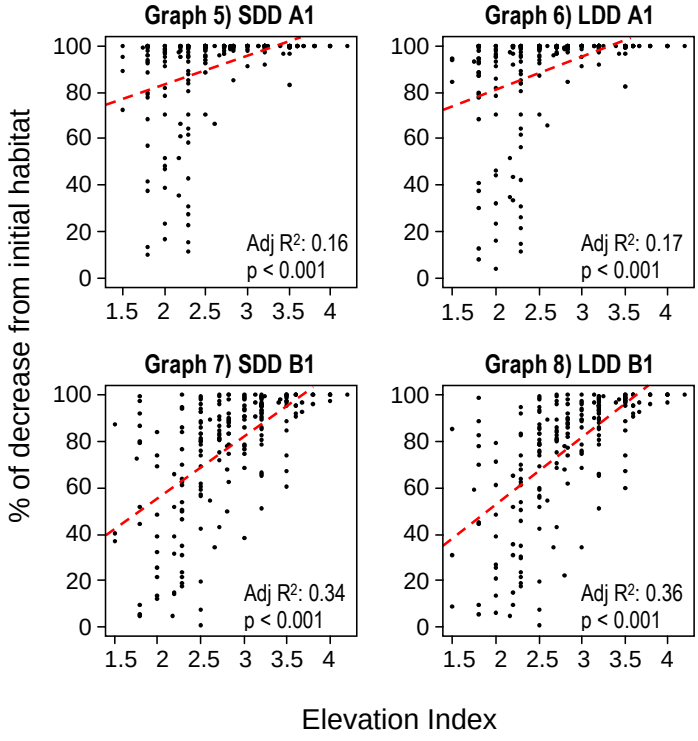
a) Decrease in predicted distribution by 2100 vs. migration index.

Climate change scenario	Dispersal scenario	Adj. R^2	Regression slope	Regression slope p-value	Nb. of species considered
A1	ND	0.017	-2.2	0.0157 *	286
	SDD Graph1	0.107	-4.6	$6.2 \cdot 10^{-8}$ ***	252
	LDD Graph2	0.070	-4.0	$1.2 \cdot 10^{-5}$ ***	251
	UD	0.056	-3.7	0.0001 ***	244
A2	ND	0.033	-3.1	0.0012 **	286
	SDD	0.051	-3.9	0.0002 ***	238
	LDD	0.088	-4.4	$2.5 \cdot 10^{-6}$ ***	231
	UD	0.038	-2.9	0.0019 **	223
B1	ND	0.069	-5.6	$3.7 \cdot 10^{-6}$ ***	286
	SDD Graph3	0.138	-7.3	$1.7 \cdot 10^{-8}$ ***	210
	LDD Graph4	0.112	-6.7	$4.5 \cdot 10^{-7}$ ***	208
	UD	0.046	-4.4	0.0014 **	197
B2	ND	0.063	-4.9	$1.1 \cdot 10^{-5}$ ***	286
	SDD	0.119	-7.1	$8.5 \cdot 10^{-8}$ ***	221
	LDD	0.098	-6.4	$1.3 \cdot 10^{-6}$ ***	218
	UD	0.031	-3.9	0.00582 **	209



b) Decrease in predicted distribution by 2100 vs. elevation Index.

Climate change scenario	Dispersal scenario	Adj. R^2	Regression slope	Regression slope p-value	Nb. of species considered
A1	ND	0.093	10.3	9.091e-08	286
	SDD Graph5	0.155	12.6	$5.5 \cdot 10^{-11}$ ***	252
	LDD Graph6	0.17	14.3	$6.1 \cdot 10^{-12}$ ***	251
	UD	0.157	14.2	$8.4 \cdot 10^{-11}$ ***	244
A2	ND	0.128	13.2	$2.6 \cdot 10^{-10}$ ***	286
	SDD	0.243	18.2	$3.9 \cdot 10^{-16}$ ***	238
	LDD	0.255	16.6	$2.2 \cdot 10^{-16}$ ***	231
	UD	0.220	15.5	$7.2 \cdot 10^{-14}$ ***	223
B1	ND	0.247	22.7	$2.2 \cdot 10^{-16}$ ***	286
	SDD Graph7	0.342	26.5	$2.2 \cdot 10^{-16}$ ***	210
	LDD Graph8	0.386	28.9	$2.2 \cdot 10^{-16}$ ***	208
	UD	0.474	33.5	$2.2 \cdot 10^{-16}$ ***	197
B2	ND	0.221	20.1	$2.2 \cdot 10^{-16}$ ***	286
	SDD	0.399	28.8	$2.2 \cdot 10^{-16}$ ***	221
	LDD	0.443	30.6	$2.2 \cdot 10^{-16}$ ***	218
	UD	0.444	32.5	$2.2 \cdot 10^{-16}$ ***	209



Chapter 3

Assessment of potential climate change impacts using fine scale data

3.1 Climate change threat to mountain plant diversity in Europe.

3.1

Climate change threat to mountain plant diversity in Europe.

Robin Engler, Christophe F. Randin, Wilfried Thuiller, Niklaus E. Zimmermann, Peter B. Pearman, Gwenaëlle Le Lay, the EUROMONT consortium and Antoine Guisan.

EUROMONT consortium: Stefan S. Dullinger, Jean-Paul Theurillat, Miguel B. Araújo, Christian Piedallu, Philippe Choler, Cécile Albert, Thomas Dirnböck, John-Arvid Grytnes, Einar Heegard, Mihai Puscas, Xavier de Lamo, Maria-Teresa Sebastià, David Nogues-Bravo, Mandar Trivedi and Angela Stanisci.

A final version of this manuscript will be submitted to *Nature.*

CONTRIBUTION TO THE PROJECT:
I assembled the data from the different participants, prepared the environmental data, performed all the modeling and analyses, analyzed the results, made the figures and wrote the initial draft of the manuscript.

Recent assessments of potential climate change impact on the distribution of animal and plant species have forecasted important extinction risks (Thomas et al. 2004, Thuiller et al. 2005). Yet, because of the coarse spatial resolution at which they were carried-out, these studies may have limited relevance for mountain systems (Randin et al. 2009, Trivedi et al. 2008). Here, using finer scale data, we assess the potential threat posed by climate change to plant species in major European mountain ranges. Using an ensemble forecasting approach (Araújo and New 2007), we model the potential distribution of 3095 species distributed over 12 study areas under current and four IPCC-based climate change scenarios for the time period 2070-2100.

Our projections suggest that, by the end of the century, the potential distribution of ~35-55% of all Alpine species, ~30-50% of all Subalpine species and ~20-45% of all Montane species could decrease by more than 80%. We also show that variability among modeling techniques can be important, and that the elevation range of the study area and the position of species along the altitudinal gradient are key factors in explaining the site-specific rates of projected species extinctions.

Comparing the standardized relative sensitivity of the different study areas to climate change suggests that vegetation from regions where warming is predicted to be low and rainfall to increase, such as the Norwegian Scandes and the Scottish Highlands, will likely be less affected by climate change than regions with greater warming and decreased precipitation, such as the Pyrenees or the Austrian Alps.

Among the different ecosystems at risk by climate change, mountain regions were early identified as particularly vulnerable (Beniston et al. 1996, Theurillat and Guisan 2001). Impacts of climate change on alpine flora were already identified over the past two decades: the first evidences of ongoing upward migration of alpine plant species in the Alps were detected by Grabherr and colleagues in 1994. Since then, additional evidence of climate change impact on mountain flora have been detected throughout several locations in Europe such as the Swiss Alps (Walther et al. 2005), the Italian Alps (Parolo and Rossi 2008), the Austrian Tyrol (Pauli et al 2007), Norway (Klanderud and Birks 2003) or Spain (Peñuelas and Boada 2003, Sanz-Elorza et al. 2003). This is of great concern because mountain ecosystems represent invaluable resources, both in terms of biodiversity and ecological services they provide (Körner 2003, Viviroli and Weingartner 2004).

Although a number of studies have already assessed the potential effects of climate change on mountain plant distributions, they have either investigated it at a broad geographic extent and coarse spatial resolution (e.g. Bakkenes et al. 2002, Thuiller et al. 2005), or locally (i.e. limited geographic extent) at fine spatial resolution (e.g. Dirnböck et al. 2003, Engler et al. 2009). In the latter case, comparing projections between local study areas is made difficult due to the different modeling settings used in the different studies. Here, we propose, for the first time, a pan-European assessment of the potential threat posed by climate change to plant species across all major European mountain ranges using high resolution data (100m or 1km grain size). We use species distribution models (Guisan and Zimmermann 2000, Guisan and Thuiller 2005) to project, under current climate and four IPCC-based climate change scenarios for the years 2070-2100, the potential distribution of 3095 species in 12 study areas distributed across all major European mountain ranges (Figure 1, Table 1). Taken together, these study areas cover approximately 130'000 km^2, representing ~5% of the European mountain surface as defined by the World Mountain Map of UNEP-WCMC (United Nations Environmental Program-World Conservation Monitoring Centre http://www.unep-wcmc.org/habitats/mountains/statistics.htm).

Table 1: Species and physical attributes of the different dataset. * = these study areas were not used in all analyses due to their small number of species.

Study Area name	Number of species	Number of plots	spatial resolution [m]	Altitudinal range [m a.s.l.]	Area [km²]
Austrian Alps	269	987	100	495 - 2265	741
Carpathians	116	968	1000	135 - 2390	38157
French Alps 1	597	2083	100	0 - 4785	57496
French Alps 2	114	274	100	1485 - 3185	63
Italian Apennines *	10	278	100	1495 - 2280	59
Norwegian Scandes	90	608	100	0 - 2445	18366
Pyrenees 1	1118	8902	1000	400 - 3135	8996
Pyrenees 2 *	5	113	100	465 - 2880	5206
Scottish Highlands	124	608	100	100 - 1200	438
Swiss inner Alps 1	265	1511	100	1505 - 4595	243
Swiss inner Alps 2	100	458	100	1875 - 3490	19
Swiss Western Alps	287	550	100	370 - 3125	704

For each species in each study area, we calibrated species distribution models using five different techniques to relate a species' presence-absence to eight topoclimatic variables (see Methods). The five resulting models were combined with an ensemble forecasting approach (Araújo and New 2007), and used to project the species' potential distribution under current (1960-1990 average) and projected future (2070-2100 average) climatic conditions. The climate change projections were based on four different socio-economic scenarios – A1FI, A2, B1, and B2 – developed by the Intergovernmental Panel on Climate Change (IPCC, Houghton et al. 2001). The A1FI climate change scenario is the most extreme, with an average warming of +5.6 °C over western and central Europe for the period 2070-2100. B1 is the mildest (+3.0 °C), and A2 (+4.5 °C) and B2 (+3.3 °C) are intermediate.

To assess a species' level of threat by climate change, we computed the proportion of potentially suitable habitat projected to be lost or gained under the different climate change scenarios. A species was classified as "extinct" or "importantly reduced" when its potentially suitable habitat was predicted to decreased by 100% or >80% respectively, as compared to its suitable habitat projected under current climatic conditions. These two thresholds were chosen because they have been used in previous studies (e.g. Thuiller et al. 2005; but see Akçakaya et al. 2006).

Modeling Results

After removing models with insufficient predictive power (see Methods), 2628 of the initial 3095 species remained for subsequent analysis. Evaluation of models calibrated for a same species was mostly very consistent across modeling techniques (Supplementary Material S4), a result which corroborates previous findings (Elith et al. 2006, Guisan et al. 2007).

In terms of spatial projections, we often found important variability between individual modeling techniques, which then propagates to the assessment of species threat levels. As a result, the variability in species extinction rates can sometimes be larger between modeling

techniques than between climate change scenarios (Figs. 1-3). Such variability between modeling techniques is not uncommon, as previously reported (e.g. Thuiller et al. 2004, Thuiller et al. 2005, Lawler et al. 2006, Pearson et al. 2006). Importantly, our implementation of ensemble forecasting yielded spatial projections having higher levels of similarity with every individual modeling technique than these among them (~80 ±20% similarity on average, see Supplementary Material S4 for full results), corroborating previous findings (Marmion et al. 2009) and supporting further implementation of ensemble forecasting, in particular the "weighted average" method that we used here, as a promising method for providing robust projections.

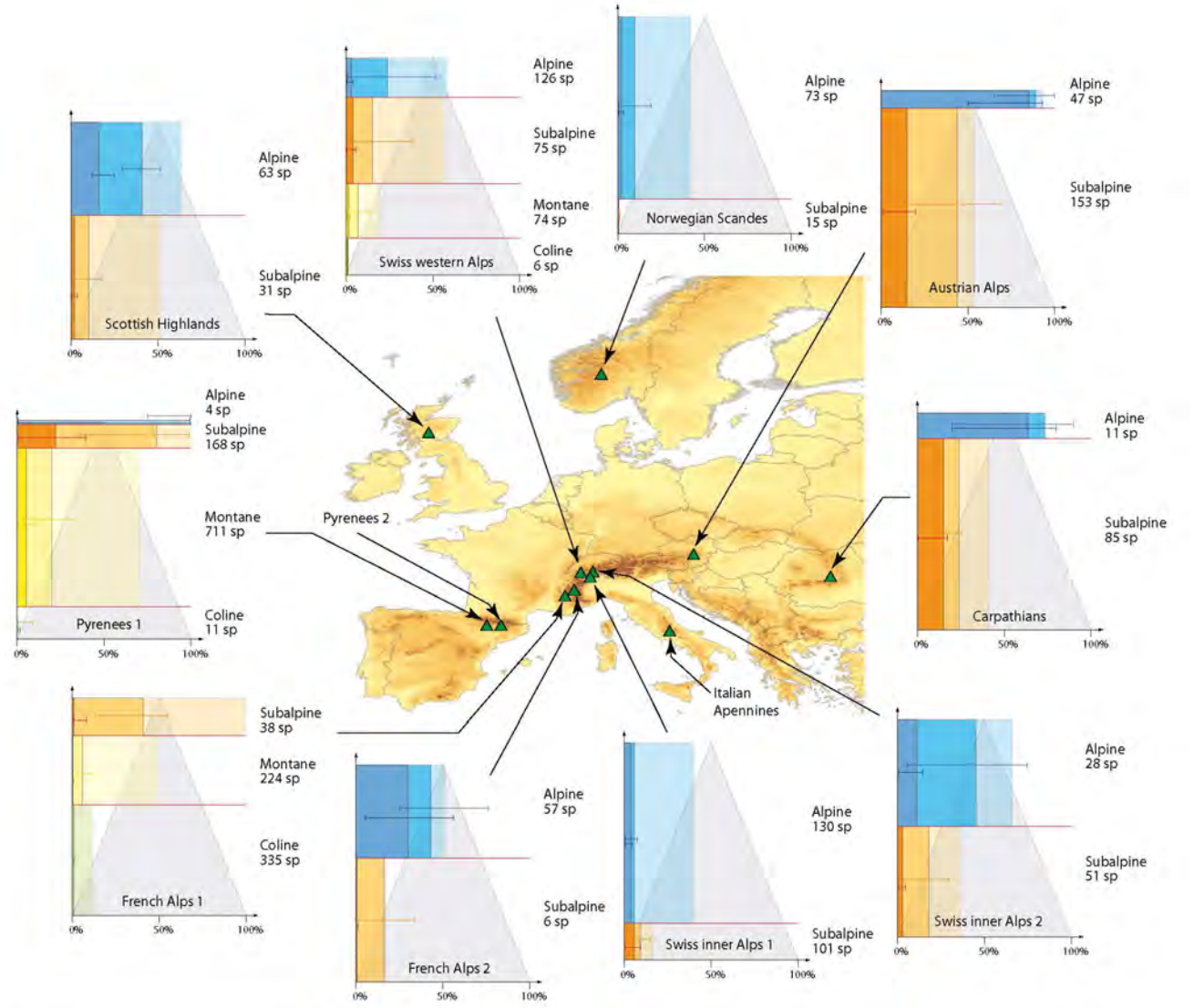
Climate change threat by the end of the 21st century

Since our study areas encompass only a portion of the European mountains, extinctions threat should be understood as local or possibly regional extinction, which does not necessarily mean total extinction in Europe. However, local extinction and species turnover might already lead to ecosystem disruptions, possibly having serious consequences on ecosystem functioning (Sala et al. 2000, Schröter et al. 2005).

Depending on the study area, the percentage of species projected to become locally extinct by 2100 varies between ~5 and 50% (mean = 25.1 ±15.9, 1sd) for the most extreme A1Fi climate change scenario, between ~1 and 45% (mean = 20.1 ±14.7, 1sd) for the A2 scenario and between ~0 and 35% (mean = 11.3 ±11.1, 1sd) under the B1 and B2 scenarios (Fig. 1 and supplementary material S5).

When considering species with projected distribution importantly reduced (i.e. decreasing by > 80%), then the number of species put at risk must roughly be doubled for A1FI and A2 climate change scenarios and tripled for B1 and B2 (see supplementary material S5 for full details).

Breaking down those numbers by altitudinal vegetation bands (Fig. 1 and supplementary material S6), there is a consistent pattern, across all studied mountain ranges, that higher elevation vegetation is forecasted to be more affected by climate change. Species from the Alpine vegetation belt appear particularly threatened, with average extinction rates across all study areas of ~40 ±29% (1sd) under A1FI and A2 and ~25 ±31% (1sd) under B1 and B2 climate change scenarios. As shown by the standard deviation values, large variation between the different study areas are observed: for instance, the study area in the Austrian Alps has projected extinction rates between ~85-90% (depending on the climate change scenario), while the Norwegian Scandes have forecasts of only ~1-10%. As will be developed later on, much of this the difference in extinction rates between study areas can be explained by the elevation range and species distribution along the elevation gradient of the study area.



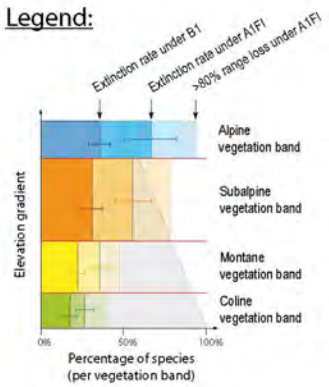


Figure 1. Percentage of species projected to become extinct (100% decrease in potential distribution) or importantly reduced (>80% decrease in potential distribution) by 2070-2100 under A1FI (most extreme) and B1 (least extreme) climate change scenarios. Results are given by study area and vegetation belt, the height of each elevation band being proportional to its relative size in the study area. Colored polygons represent the average value obtained from the two "weighted average" projection based on respectively AUC and TSS thresholds (see Methods). Errors bars indicate the range of values observed in 80% of the individual models that yielded results closest to the average value, which is here equivalent of showing the range of results yielded by all but the two most outlier models. Medium shaded polygons indicate projected extinction rates under A1FI and dark shaded extinction rates under B1. Light shaded polygons indicate the percentage of species whose potential distribution is projected to decrease > 80% under A1FI, no error bar is given for this value.

Pooling all species together by vegetation band reveals that, depending on the considered climate change scenario and across all study areas, ~35-55% of all Alpine species, ~30-50% of all Subalpine species and ~20-45% of all Montane species are predicted to become locally "importantly reduced" by the end of the century, (Fig. 2). While the majority of Alpine, Subalpine and Montane species are expected to lose an important part of their potentially suitable range due to climate change, species belonging to the Coline vegetation band are, on the other hand, mostly predicted to increase in potential distribution.

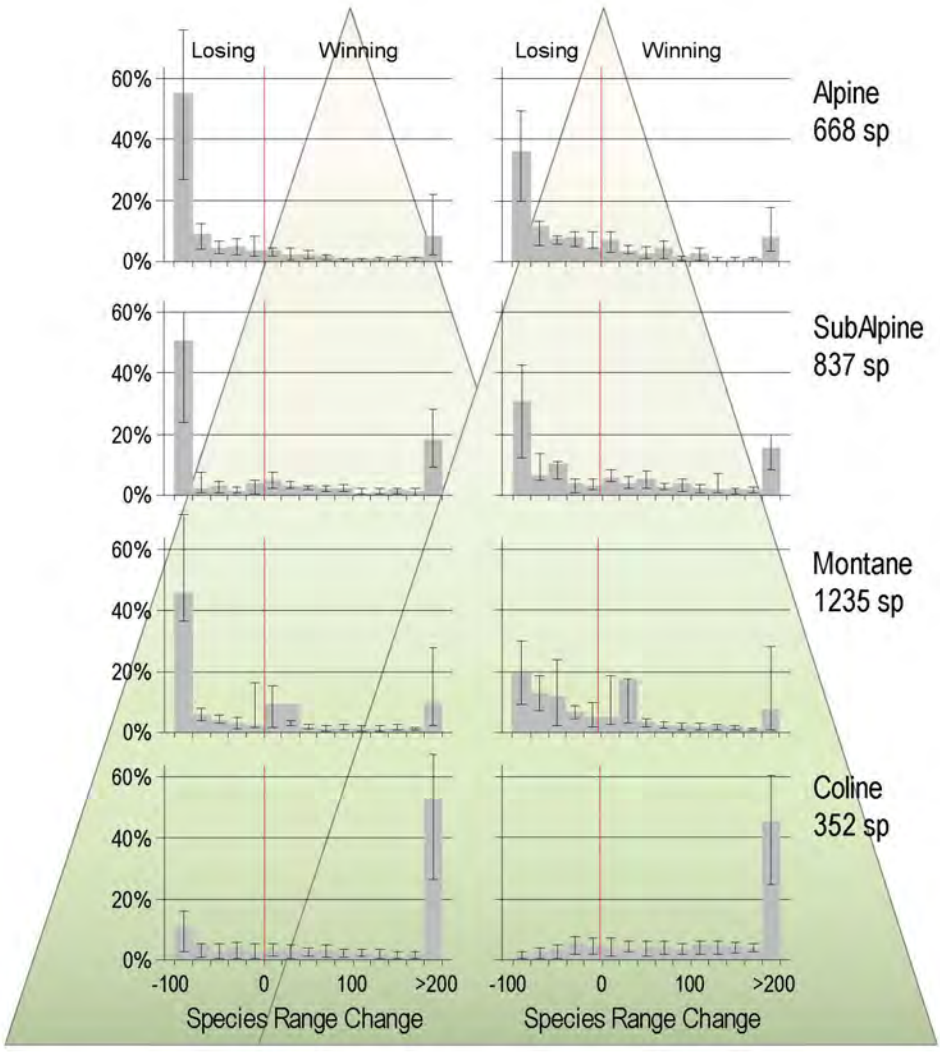


Figure 2. Average distribution of Species Range Change (SRC) per vegetation belt for all species of all study areas under A1FI (left panels) and B1 (right panels) climate change scenarios. Histogram bars represent the average value obtained from the AUC and TSS weighted average projections. Errors bars indicate the range of values observed in 80% of the individual models that yielded results closest to the average value. This is equivalent to showing the range of results yielded by all but the two most outlier models. The vertical red bar indicates the separation between species projected to decrease in distribution ("losing") or increase in distribution ("winning").

Compared to Thuiller et al. (2005), who projected ~60% species extinctions for mountain areas, our results forecast a lower magnitude of threat to mountain flora, and even more so as we project distributions for the 2070-2100 whereas Thuiller et al. (2005) projected for the 2051-2085. This likely results from coarse scale models overestimating threat levels from climate change due to their inability to capture local, fine scale, microclimatic features (Randin et al. 2009).

A number of other factors may also further mitigate the species threat levels projected by our models. For instance, genetic adaptation and phenotypic plasticity may buffer the effects of climate change (Theurillat and Guisan 2001). Evidence from Quaternary (~2.5 million years BCE to present) suggests surprisingly little species extinctions although climate change of similar magnitude than what projected for the near future has sometime occurred during this period (Botkin et al. 2007).

Climate change threat by endemism and growth form

Comparing the projections of species range change (i.e. proportion of potential habitat gain or loss) by 2100 between endemic (94 sp.) and non-endemic (1277 sp.; 1257 sp. could not be classified) species showed no particular pattern in terms of threat level (see supplementary material S7). Endemic species did thus not appear more vulnerable than non-endemic ones. This is expected because of the local/regional extent of each study area. Yet, while endemic and non-endemic species are projected to have similar local extinction risk, their level of threat when considering a larger geographical scale might well be different, since, for an endemic species, local/regional extinction could mean complete extinction, whereas populations of more ubiquitous species might still persist in other locations within or outside Europe.

Comparing the proportion of potential habitat gain or loss between species growth forms (see Methods) reveals that chameaphyte cushion plants are projected to undergo especially high surface loss (~75% and ~55% of species are projected to decrease in distribution by >80% under respectively the A1FI and B1 scenarios). All other species growth forms appear fairly equally threatened by climate change (~45% – A1FI – and ~25% – B1 – of species are projected to lose >80% of distribution) except for trees that are projected to be more resilient to climate change (only ~35% – A1FI – and ~15% – B1 – of species are predicted lose 80% or more of their distribution). Such results are consistent with the fact that cushion type plants are typically those occupying the highest elevations niches (Körner 2003), while trees occupy the lower part of the altitudinal gradient and have thus the opportunity to migrate uphill to find new suitable conditions.

However, as our analysis did not account for possible new competing species invading the lower elevation areas nor for dispersal limitations, threat level to some species and especially those inhabiting lower elevations could be higher than projected here.

Relative sensitivity of individual European mountain ranges to climate change

Since the study areas and datasets representing each mountain range have different size, elevation gradient and species pools, rates of species projected to become locally extinct or importantly reduced cannot be compared directly between them. The two most important bias factors that could be identified to introduce "artificial" variation in species threat levels

between study areas are (i) elevation range, i.e. study areas with a larger elevation range are expected to have less species extinctions because they offer more room for species to migrate upwards, and (ii) position of species along the altitudinal gradient of the study area, i.e. species located towards the bottom of the study area are expected to be less threatened because they are provided with the opportunity to migrate upwards. We assessed the importance of these two factors by regression and developed standardized threat levels comparable between areas (see Methods).

Considered together, these two factors explained on average ~65% and ~40% of the variation in the rate of species projected to become extinct (100% potential distribution loss) or importantly reduced (>80% decrease in potential distribution) respectively. Both factors did almost always explain a highly significant proportion of variance in the data (p-value < 0.01, anova Chi-square test). This result confirms that a large proportion of the variance in species threat levels between study areas is explained by the study area structure and the distribution of the investigated species along the altitudinal gradient.

By comparing the regression residuals of our different study areas, we were able to derive the relative climate sensitivity of each mountain range and compare them. Although there is some variation depending on the climate change scenario, the "threat threshold" and the employed modeling technique, overall, the following trend emerges (Fig. 3): The study

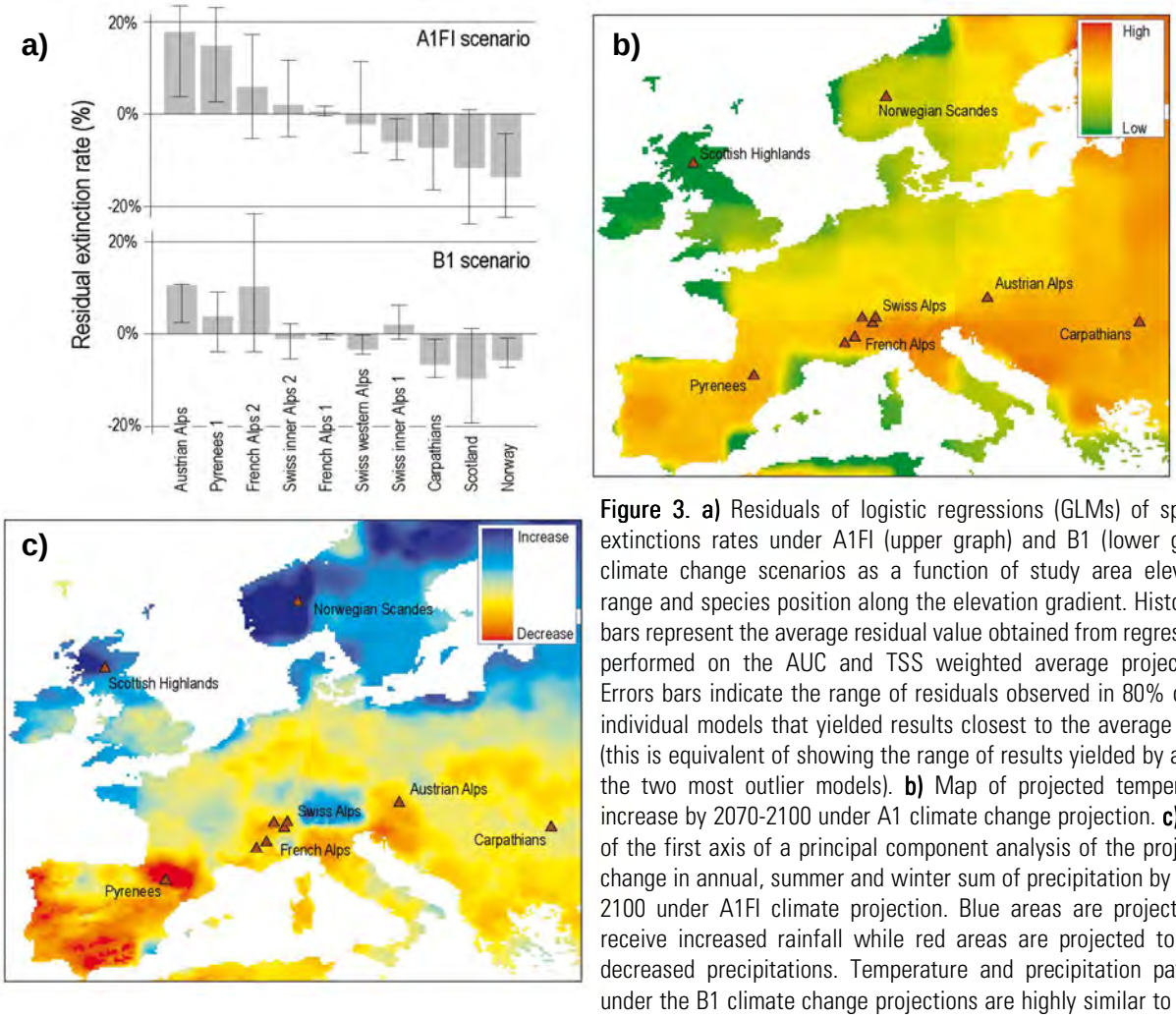


Figure 3. a) Residuals of logistic regressions (GLMs) of species extinctions rates under A1FI (upper graph) and B1 (lower graph) climate change scenarios as a function of study area elevation range and species position along the elevation gradient. Histogram bars represent the average residual value obtained from regressions performed on the AUC and TSS weighted average projections. Errors bars indicate the range of residuals observed in 80% of the individual models that yielded results closest to the average value (this is equivalent of showing the range of results yielded by all but the two most outlier models). **b)** Map of projected temperature increase by 2070-2100 under A1 climate change projection. **c)** Map of the first axis of a principal component analysis of the projected change in annual, summer and winter sum of precipitation by 2070-2100 under A1FI climate projection. Blue areas are projected to receive increased rainfall while red areas are projected to have decreased precipitations. Temperature and precipitation patterns under the B1 climate change projections are highly similar to those shown for the A1FI projections.

areas from the Austrian Alps and Pyrenees are projected to be more sensitive to climate change than average, study areas from the Swiss and French Alps are average and the study area from the Scottish Highlands, the Norwegian Scandes and the Carpathians appeared less sensitive than average. A similar pattern was found by Thuiller et al. (2005), who also projected the Scottish Highlands, the Norwegian Scandes and the Carpathians to be less sensitive than other European mountain ranges for the 2051-2080 time period. Future studies could estimate how sensitive their study area is in respect to other European mountain ranges by comparing their results against our regression models (see supplementary material S8).

Explanations to this result partly lie in the fact that Scottish Highlands and Norwegian Scandes have the smallest mean annual temperature increase (+3.65°C and +5°C respectively) under the A1FI climate change scenario and the smallest and third smallest increase under B1 (+2.1 and +2.8°C respectively). Looking at projected change in precipitation regimes, the Norwegian Scandes and the Scottish highlands also contrast with the other study areas in that they are projected to receive increased rather than decreased rainfall in the future (Fig. 3c, blue areas). It is also interesting to note that, for these study areas, rainfall shows no correlation with elevation (*Pearson r* = ~0.1), which contrasts with other study areas where the correlation between elevation and rainfall is stronger (*mean Pearson r*= 0.65 ±0.4, 1sd). Looking at the study areas that are projected to be more sensitive to climate change than average (Austrian Alps and Pyrenees), their projections predict a larger than average temperature increase (Fig. 3b) and an important decrease in rainfall (Fig. 3c, red areas).

This suggests that study areas with a projected increase in precipitation and/or a weak correlation between precipitation and elevation could lead models to predict the flora in those study areas as less threatened by climate change. Conversely, a projected decrease in rainfall and a stronger correlation of precipitation with elevation could lead species distribution models to project higher rates of plant species extinctions because those areas that were suitable no longer exits (i.e. the model requires the species to track wet conditions that no longer exist in the climate change projections, just like what happens with an increase in temperature). Our results are concordant with findings from Thuiller et al. (2005) who also obtained higher rates of species extinctions in regions projected to have higher temperature warming and decrease in rainfall.

An additional element that could explain the difference in projected sensitivity to climate change between study areas is the influence of a given environmental variable on a species' model. Here we measured the influence of each variable in each model (see Methods) and compared it between study areas and between species projected to respectively decrease or increase in distribution under climate change.

For both species projected to decrease and increase in distribution, mean annual temperature was, on average, the most influent variable in the models. However, the frequency with which mean annual temperature was selected as most or second most important variable for species projected to undergo a decrease in potential distribution was significantly higher than that of species projected to increase in distribution (binomial proportions test, p-value < 10^{-12}, N=576). This was not the case for any other variable except summer moisture index. Furthermore, mean annual temperature was also found to be more frequently the most influential variable in species distribution models for the Austrian Alps and the Pyrenees 1 datasets, both of which are predicted as more sensitive to climate change than average.

This suggests that models for which the temperature variable is a major driver will predict reduction in potential distribution more so than others. Given the very strong correlation between temperature and elevation, this result is expected: models relying heavily on temperature will project their species' suitable habitat to shift upwards, which mostly translates into a reduction or complete disappearance of their habitat.

Taken together, our results suggests that while there is no one factor that can explain alone why some mountain regions are projected to be more, or less, sensitive to climate change than others, the factors influencing this sensitivity include: (i) regions with lower warming and increased rainfall appear less affected than those with higher warming and reduction in rainfall, (ii) regions where the environmental variables affected by climate change are not correlated with elevation appear as less sensitive to climate change, and (iii) regions where species distribution models are more frequently driven by temperature-based variables appear as more sensitive to climate change. Although our study already provided a comprehensive assessment of climatic threat to mountain floras in Europe, our findings ought to be verified for other mountain ranges in Europe, and elsewhere in the world.

Methods

Study areas and species datasets

For this study, we assembled vegetation survey data from 12 study areas covering all major mountain ranges from western and central Europe (Table 1, Fig. 1). To prevent loss of robustness in calibration of distribution models, species with less than 20 occurrence records within a dataset were discarded, leaving a total of 3'095 species, some of which were redundant between study areas. Spatial accuracy of the vegetation plots was of 100 m or better for 10 study areas and of 1 km or better for the remaining two datasets (Carpathians and Pyrenees 1 datasets, see Table 1). The spatial accuracy of records dictated the resolution at which spatial projection and analysis could be carried out for each study area.

Environmental variables

We employed eight topo-climatic variables known to represent physiological requirements of plant species (Pearson et al. 2002, Körner 2003): mean annual temperature, mean temperature of the coldest month, annual, summer (July-September) and winter (January-March) sum of precipitations, annual, summer (July-September) and winter (January-March) moisture index (see supplementary material S1 for details on variable computation). The moisture index was calculated as the difference between precipitation and potential evapotranspiration, expressing the amount of soil water that is potentially available at a site. Evapotranspiration was itself computed by accounting for elevation, slope, aspect, shadowing, cloud cover, latitude and longitude.

All variables were derived from long-term (1960-1990 average) monthly means for average temperature, sum of precipitations and average cloudiness, combined with data from digital elevation models. Variables were prepared identically for all study area and at a spatial resolution matching the one of the records of the species datasets, i.e. either 100m or 1km pixel size (see Table 1). Full details of environmental variable preparation are given in supplementary material S1.

Climate change scenarios

We used four different climate projections for the 2070-2100 time period developed by the UK Hadley Center for Climate Prediction and Research (Mitchell et al. 2003, Mitchell and Jones 2005). These were derived from a global circulation model (HadCM3; Carson 1999), and are based on four different socio-economic scenarios A1FI, A2, B1, and B2 developed by the IPCC (Intergovernmental Panel on Climate Change; Houghton et al. 2001). These climate change projections were available in the form of 10' (~15 km in Europe) grids of monthly average temperatures, monthly sums of precipitations and monthly average cloud cover. Each of the four climate change projections was averaged over the 2070-2100 time period and downscaled using bilinear interpolation to match the spatial resolution of the considered study area (full details of how climate change projections were implemented are given in supplementary material S1).

With an average increase of +5.6 °C over western and central Europe for the period 2070–2100, the A1FI climate change scenario is the most extreme. B1 is mildest (+3.0 °C), and A2 (4.5 °C) and B2 (3.3 °C) are intermediate.

Species distribution modeling

For each species and each dataset, we related a species' presence-absence to the previously described eight environmental variables using five different modeling techniques: generalized linear models (GLM; McCullagh and Nelder 1989), generalized additive models (GAM; Hastie and Tibshirani 1986), generalized boosted models (GBM; Ridgeway 1999), random forest (RF; Breiman 2001) and multivariate adaptive regression splines (MARS; Friedman 1991). All models were calibrated using the Biomod framework (Thuiller et al. 2009) in R (R Development Core Team 2009).

GLMs and GAMs were calibrated using a binomial distribution and a logistic link function. A bi-directional stepwise procedure in was used for predictor selection, based on the Akaike information criterion (AIC; Akaike 1973). Up to second-order polynomials (linear and quadratic terms) were allowed for each predictor.

The predictive power of each model was evaluated through a data-splitting procedure (see Thuiller et al. 2009 for detailed explanations): a model was trained on 70% of the data before being evaluated on the remaining 30%. This splitting was done while maintaining the original prevalence of a species' presences and absences in the data. The agreement between predicted and observed values was then evaluated using two different measures: the area under the relative operating characteristic curve (AUC; Hanley and McNeil 1982) and the True Skill Statistic (TSS; Allouche et al. 2006). This data-splitting procedure was repeated 50 times and the evaluation values averaged. While model evaluation was done using the above-mentioned data-splitting procedure, the final models used for spatial projections were calibrated using 100% of the data for a species in a given study area, thereby allowing to take advantage of the full available data.

Finally we also measured the influence of each environmental variable in each model by measuring the impact that a randomization in the variable had on the model's projections (see Thuiller et al. 2009 for detailed explanations on this method). Randomization was repeated 100 times for each variable and the results averaged.

Spatial projections

For each species and each modeling technique, projections were carried-out over the entire study area under current climatic conditions (i.e. 1960-1990 average) and the four projections of future climate (i.e. A1FI, A2, B1 and B2) averaged over the year 2070-2100. Following an "ensemble forecasting" approach (Araújo and New 2007) that aims at providing more robust projections by reducing the variability found between individual modeling techniques, we also computed two "weighted average" projections. These were obtained by averaging the projections from the five individual modeling techniques that were weighted respectively by their AUC or TSS scores (details on calculations are given in supplementary material S2), and will thereafter be referred to as "AUC weighted average" and "TSS weighted average". This weighted average method was shown to be particularly robust (Marmion et al. 2009). To avoid basing projection on poorly calibrated models, the weighted average projections were computed using only those models with AUC and TSS values respectively > 0.7 and > 0.4.

To produce maps of potentially suitable habitat, we reclassified the probabilistic projections (in the range [0:1]) into binary values, representing either suitable or unsuitable habitat. This conversion from probabilistic to binary values requires the selection of a threshold above which a pixel is reclassified as potentially suitable and unsuitable below. Each probabilistic map was here reclassified using two different thresholds: (i) maximizing jointly the percentage of presences and absences correctly predicted, i.e. the probability where sensitivity = specificity (Liu et al. 2005). This threshold is here referred to as "AUC threshold". (ii) maximizing the value of the TSS evaluation value, here referred to as "TSS threshold". Each individual model thus yielded two different projections of potential current and future distribution for a given species in a given study area.

Altogether, we generated 12 different projections for each species in each study area: two projections for each of the five individual modeling techniques reclassified with either their AUC or TSS threshold, one for the "AUC weighted average" reclassified with its AUC threshold and one for the "TSS weighted average" reclassified with its TSS threshold.

We subsequently applied a mask representing heavily anthropized areas (e.g. urban areas), water bodies and glaciers to avoid projections at locations that are unsuitable regardless of climate. Finally, projections were further restricted to land cover classes (forests, grasslands, rocks/scree) in which a species was observed at least once (e.g. a grassland species could not be projected to occur in forested areas).

Species were assumed to have unlimited dispersal capability, thus occupying all pixels projected to become potentially suitable as a result of climate change. This assumption might be conservative (i.e. optimistic), but was shown to provide good estimates of species threat levels for mountain areas (Engler et al. 2009).

Data analysis

To avoid working with data from poorly calibrated models, only projections from models with AUC > 0.7 and TSS > 0.4 were considered in all analyses. These values represent what is generally considered as the minimum evaluation scores for a model to be "useful" (Araujo et al. 2005, adapted from Swets 1988).

In order to analyze the data by altitudinal zones, each study area was divided into one or more of the four vegetation belts: Alpine, Subalpine, Montane and Coline. The vegetation belts were defined separately for each study area based on local knowledge (see

supplementary material S3 for details). Each species in each study area was in turn associated to the vegetation belt where the majority of its field observations fell into.

When possible, each species was also classified according to its endemism status (either locally/regionally endemic or non-endemic) and its growth form (therophyte, geophyte, hemicryptophyte forb or graminoid, chameaphyte cushion or dwarf shrub, phanerophyte shrub or tree).

To assess the effect of the elevation range and the position of species along the altitudinal gradient on the projected species extinctions, we fitted binomial logistic regressions (generalized linear models) using a study area's species threat level as response variable and elevation range and altitudinal position of species as explanatory variables (see supplementary material S8 for details). We then used the regression residuals as an indication of relative threat level that is comparable between study areas.

References

Akçakaya H.R., Butchart S.H.M., Mace G.M., Stuart S.N. and Hilton-Taylor C., 2006. Use and misuse of the IUCN Red List Criteria in projecting climate change impacts on biodiversity. *Global Change Biology*, **12**, 2037-2043.

Araújo M.B., Pearson R.G., Thuiller W. and Erhard M., 2005. Validation of species - climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.

Araujo M.B. and New M., 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**(1), 42-47.

Bakkenes M., Alkemade J.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**(4), 390-407.

Beniston M., Fox D.G., Adhikary S., Andressen R., Guisan A., Holten J.I., Innes J., Maitima J., Price M.F. and Tessier L., 1996. Impacts of climate change on mountain regions. Chapter 5. *in* IPCC, editor. Intergovernmental panel on climate change. Second assessment report. Cambridge University Press, Cambridge.

Botkin D.B., Saxe H., Araujo M.B., Betts R., Bradshaw R.H.W., Cedhagen T., Chesson P., Dawson T.P., Etterson J.R., Faith D.P., Ferrier S., Guisan A., Hansen A.S., Hilbert D.W., Loehle C., Margules C., New M., Sobel M.J. and Stockwell D.R.B., 2007. Forecasting the effects of global warming on biodiversity. *Bioscience*, **57**, 227-236.

Dirnböck T., Dullinger S. and Grabherr G., 2003. A regional impact assessment of climate and land-use change on alpine vegetation. *Journal of Biogeography*, **30**(3), 401-417.

Elith J., Graham C.H., Anderson R.P., Dudik M., Ferrier S., Guisan A., Hijmans R.J., Huettmann F., Leathwick J.R., Lehmann A., Li J., Lohmann L.G., Loiselle B.A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J.M., Peterson A.T., Phillips S.J., Richardson K., Scachetti-Pereira R., Schapire R.E., Soberon J., Williams S., Wisz M.S. and Zimmermann N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**(2), 129-151.

Engler R., Randin C.F., Vittoz P., Czáka T., Beniston M., Zimmermann N.E. and Guisan A., 2009. Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? *Ecography*, **32**, 34-45.

Grabherr G., Gottfried M. and Pauli H., 1994. Climate Effects on Mountain Plants. *Nature*, **369**(6480), 448-448.

Guisan A. and Zimmermann N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**(2-3), 147-186.

Guisan A. and Thuiller W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Guisan A., Zimmermann N.E., Elith J., Graham C., Phillips S. and Peterson A.T., 2007. What matters for predicting spatial distributions of trees: techniques, data, or species' characteristics? *Ecological Monographs*, **77**(4), 615-630.

Houghton J.T., Ding Y. and Griggs D.J. (eds.). 2001. Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. - Cambridge University Press.

Klanderud K. and Birks H.J.B., 2003. Recent increases in species richness and shifts in altitudinal distributions of Norwegian mountain plants. *The Holocene*, **13**, 1-6.

Körner C., 2003. Alpine plant life. 2nd Ed. *Springer*.

Lawler J.J., White D., Neilson R.P. and Blaustein A.R., 2006. Predicting climate-induced range shifts: model differences and model reliability. *Global Change Biology*, **12**(8), 1568-1584.

Marmion M., Parviainen M., Luoto M., Heikkinen R.K. and Thuiller W., 2009. Evaluation of consensus methods in predictive species distribution modeling. *Diversity and Distributions*, **15**, 59-69.

Parolo G. and Rossi G., 2008. Upward migration of vascular plants following a climate warming trend in the Alps. *Basic and Applied Ecology*, **9**, 100-107.

Pauli H., Gottfried M., Reiter K., Klettner C. and Grabherr G., 2007. Signals of range expansions and contractions of vascular plants in the high Alps: observations (1994–2004) at the GLORIA master site Schrankogel, Tyrol, Austria. *Global Change Biology*, **13**(1), 147-156.

Pearson R.G., Thuiller W., Araujo M.B., Martinez-Meyer E., Brotons L., McClean C., Miles L., Segurado P., Dawson T.P. and Lees D.C., 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**(10), 1704-1711.

Peñuelas J. and Boada M., 2003. A global change-induced biome shift in the Montseny mountains (NE Spain). *Global Change Biology*, **9**, 131-140.

Randin C.F., Engler R., Normand S., Zappa M., Zimmermann N.E., Pearman P.B., Vittoz P., Thuiller W. and Guisan A., 2009. Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology*, **15**, 1557-1569.

Sala O.E., Chapin F.S., 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. and Wall D.H., 2000. Biodiversity - Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770-1774.

Sanz-Elorza M., Dana E.D., Gonzalez A. and E. S., 2003. Changes in the high-mountain vegetation of the central Iberian Peninsula as a probable sign of global warming. *Annals of Botany*, **92**, 1-8.

Schröter D., Cramer W., Leemans R., Prentice I.C., Araujo M.B., Arnell N.W., Bondeau A., Bugmann H., Carter T.R., Gracia C.A., de la Vega-Leinert A.C., Erhard M., Ewert F., Glendining M., House J.I., Kankaanpaa S., Klein R.J.T., Lavorel S., Lindner M., Metzger M.J., Meyer J., Mitchell T.D., Reginster I., Rounsevell M., Sabate S., Sitch S., Smith B., Smith J., Smith P., Sykes M.T., Thonicke K., Thuiller W., Tuck G., Zaehle S. and Zierl B., 2005. Ecosystem service supply and vulnerability to global change in Europe. *Science*, **310**, 1333-1337.

Theurillat J.P. and Guisan A., 2001. Potential impact of climate change on vegetation in the European Alps: A review. *Climatic Change*, **50**(1-2), 77-109.

Thomas C.D., Cameron A., Green R.E., Bakkenes M., Beaumont L.J., Collingham Y.C., Erasmus B.F.N., Ferreira de Siqueira M., Grainger A., Hannah L., Hughes L., Huntley B., Van Jaarsveld A.S., Midgley

G.F., Miles L., Ortega-Huerta M.A., Peterson A.T., Phillips O.L. and Williams S.E., 2004. Extinction risk from climate change. *Nature*, **427**(6970), 145-147.

Swets J.A., 1988. Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.

Thuiller W., Lavorel S., Araujo M.B., Sykes M.T. and Prentice I.C., 2005. Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(23), 8245-8250.

Thuiller W., Araújo M.B., Pearson R.G., Whittaker R.J., Brotons L. and Lavorel S., 2004. Uncertainty in predictions of extinction risk. *Nature*, **430**, 34.

Trivedi M.R., Berry P.M., Morecroft M.D. and Dawson T.P., 2008. Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. *Global Change Biology*, **14**(5), 1089-1103.

Viviroli D. and Weingartner R., 2004. Hydrological significance of mountains: from regional to global scale. Hydrological Earth Systems. *Science*, **8**(6), 1016–1029.

Walther G.-R., Beissner S. and Burga C.A., 2005. Trends in the upward shift of alpine plants. *Journal of Vegetation Science*, **16**(5), 541-548.

Supplementary Material / Supporting Information

Supplementary material S1: Environmental variable preparation methodology.

For each of the 12 study areas, eight different topo-climatic environmental variables were prepared at a spatial resolution (pixel size) matching that of the dataset's vegetation plot records (i.e., 1 km for the Carpathians and the Pyrenees 1 datasets and 100 m for all other datasets, Table S1.1). Each variable was prepared under current climatic conditions (1960-1990 average) and four different climate change projections for the 2070-2100 average future time period.

Table S1.1: Environmental variable resolution (first column) and physical attributes of the different datasets.

Study Area name	spatial resolution [m]	Area [km^2]	Pixel number	Altitudinal range [m a.s.l.]
Austrian Alps	100	741	74'100	495 - 2265
Carpathians	1000	38'157	38'157	135 - 2390
French Alps 1	100	57'496	5'749'600	0 - 4785
French Alps 2	100	63	6'300	1485 - 3185
Italian Apennines	100	59	5'900	1495 - 2280
Norwegian Scandes	100	18'366	1'836'600	0 - 2445
Pyrenees 1	1000	8'996	8'996	400 - 3135
Pyrenees 2	100	5'206	520'600	465 - 2880
Scottish Highlands	100	438	43'800	100 - 1200
Swiss inner Alps 1	100	243	24'300	1505 - 4595
Swiss inner Alps 2	100	19	1'900	1875 - 3490
Swiss Western Alps	100	704	70'400	370 - 3125

Environmental variables under current climatic conditions

All environmental variables were derived from the following basic variables: monthly average temperatures [°C], monthly sums of precipitations [mm], monthly average cloudiness [percent cover] and digital elevation model [m]. All temperature, precipitation and cloudiness data were averages for the period 1960-1990 or 1961-1990.

Digital Elevation Models (DEM) data were generally available from national topographic offices at a native resolution of 100 m or less, in which case the data was aggregated to a 100 m resolution. For study areas were no native fine resolution DEM was available, data were extracted from the 3-arc-second (~90 m at equator) Shuttle Radar Topography Mission DEM (SRTM; USGS 2003) and resampled to 100 m pixel size.

Monthly temperature and precipitation data were generally also available at fine scale (i.e. 100 meters or less) from different national meteorological or topographic offices. When no native fine resolution data was available for average temperature and/or sum of

precipitations, 1 km resolution WorldClim data (Hijmans et al. 2005) was used and downscaled to 100 meters. Average temperatures were downscaled using the *tave* ArcGIS program (see Zimmermann and Kienast, 1999), which computes a local temperature lapse rate from the temperature data and a DEM in order to interpolate the data while accounting for a pixel's elevation. Sum of precipitations data were downscaled using simple bilinear interpolation.

Data for monthly average cloud cover with a 10 minute (~15 km in Europe) resolution were obtained from the UK Hadley Center for Climate Prediction and Research (Mitchell et al. 2003). These maps were downscaled to either 100 m or 1 km resolution (whatever was matching the dataset's resolution, see Table S1.1) using bilinear interpolation.

Using the above described data as basis for computation, the following eight topo-climatic variables were derived (Table S1.2):

Table S1.2: Environmental variables derived from the basic variables and used for species distribution modeling.

Variable	Unit	Description
mean annual temperature	[°C]	Average daily temperature over the entire year
mean temperature of coldest month	[°C]	mean temperature of coldest month
Annual sum of precipitations	[mm × year^{-1}]	Sum of precipitations over the entire year
Summer sum of precipitations	[mm × 3 month^{-1}]	Sum of precipitations from July to September
Winter sum of precipitations	[mm × 3 month^{-1}]	Sum of precipitations from January to March
Annual moisture index	[mm × year^{-1}]	Sum of potentially available water over the entire year
Summer moisture index	[mm·3 month^{-1}]	Sum of potentially available water from July to September
Winter moisture index	[mm × 3 month^{-1}]	Sum of potentially available water from January to March

The details of each variable's computation are given here-below:

Mean annual temperature:

Let $MTave_i$ be the average temperature for month i and $NDays_i$ the number of days in that same month. The mean annual temperature was calculated as follows (Eq. S1.1):

$$Mean\ Annual\ Temperature = \frac{1}{365} \times \sum_{i=1}^{12} MTave_i \times NDays_i \qquad \text{Eq. S1.1}$$

Mean temperature of the coldest month:

The mean temperature of the coldest month was simply calculated by selecting, for each pixel, the months during which the mean annual temperature (*MTave*) was minimal.

$$Mean\ Temperature\ of\ Coldest\ Month = Min(i \in 1-12)\ MTave_i \qquad \text{Eq. S1.2}$$

Annual, summer and winter sum of precipitations:

Let $MPrec_i$ be the sum of precipitations for month *i*. Annual, summer and winter sums of precipitations were computed as follows (Eq. S1.3 – S1.5):

$$Annual\ Sum\ of\ \mathrm{Pr}eciptiations = \sum_{i=1}^{12} M\ \mathrm{Pr}ec_i \qquad \text{Eq. S1.3}$$

$$Summer\ Sum\ of\ \mathrm{Pr}eciptiations = \sum_{i=7}^{9} M\ \mathrm{Pr}ec_i \qquad \text{Eq. S1.4}$$

$$W\,\mathrm{int}er\ Sum\ of\ \mathrm{Pr}eciptiations = \sum_{i=1}^{3} M\ \mathrm{Pr}ec_i \qquad \text{Eq. S1.5}$$

Annual, summer and winter moisture index:

The moisture index expresses the amount of soil water that is potentially available at a given site (pixel). It is calculated as the difference between precipitation and potential evapotranspiration, the later variable being a function of solar radiations and average temperature. Since solar radiation was not a readily available variable, we first had to calculate it by accounting for local topography, geographical position and cloudiness.

For each study area, monthly average of daily sum of solar radiation was computed based on a digital elevation model and the monthly average cloud cover data. Calculation was carried-out using the "Helios" program (Piedallu and Gegout 2007), which accounts for elevation, slope, aspect, shadowing, cloud cover, latitude and longitude.

Solar radiations were then combined with average temperature to derive potential evapotranspiration (*PET*) using the Jensen-Haise empirical formula (Jensen and Haise 1963; Eq. S1.5):

$$PET\ [\mathrm{mm/day}] = \left(4 \times \left(0.0239001 \times \mathrm{Rs} + 50\right)\right) \times \left({\left(Ta \times 30\right)} \Big/ {\left(Ta + 15\right)} \right) \qquad \text{Eq. S1.5}$$

Where *Rs* is the daily global radiation (monthly average) in [kJ × m^{-2} × day^{-1}] and *Ta* the monthly average temperature [°C]. Eq. S1.5 allowed deriving a daily potential evapotranspiration value which could then be used to compute a monthly sum.

With $MPrec_i$ the sum of precipitation for month *i*, and $MPET_i$ the sum of potential evapotranspiration for month *i*, annual, summer and winter moisture indexes could be computed as follows (Eq. S1.6 – S1.8):

$$Annual\ Moisture\ Index = \sum_{i=1}^{12} M\ \Pr ec_i - MPET_i \qquad \text{Eq. S1.6}$$

$$Summer\ Moisture\ Index = \sum_{i=7}^{9} M\ \Pr ec_i - MPET_i \qquad \text{Eq. S1.7}$$

$$W\ \text{int}\ er\ Moisture\ Index = \sum_{i=1}^{3} M\ \Pr ec_i - MPET_i \qquad \text{Eq. S1.8}$$

Note that, in Eq. S1.6 - S1.8, whenever monthly potential evapotranspiration was greater than the monthly sum of precipitations, zero was added to the summation (i.e. values of $MPerc_i - MPET_i < 0$ were set to zero before being added to the sum). This makes sense as no more water can evaporate than what is available.

Environmental variables under projected future climatic conditions

We used four different climate projections for the 2070-2100 time period developed by the UK Hadley Center for Climate Prediction and Research (Mitchell et al. 2003, Mitchell and Jones 2005). These were derived from a global circulation model (HadCM3; Carson 1999), and are based on four different socio-economic scenarios A1FI, A2, B1, and B2 developed by the IPCC (Intergovernmental Panel on Climate Change; Houghton et al. 2001). These climate change projections were available in the form of 10' (~15 km in Europe) grids of monthly average temperatures, monthly sums of precipitations and monthly average cloud cover.

For each month and each climate projection scenario, we averaged each of these three variables over the 2070-2100 time period. This, in turn, allowed computing monthly anomalies (i.e., difference) for each variable between the average of 1960-1990 and the projected average of 2070-2100 under each climate change scenario.

These anomalies were then downscaled to 100 m or 1 km resolution (whatever was matching the dataset's resolution, see Table S1.1) using bilinear interpolation before being added to their corresponding variable under current climatic condition. Thereafter, we calculated the derived environmental predictors under each of the four climatic projections (A1FI, A2, B1 and B2) using the same methodology as when deriving those environmental variables under current climatic conditions (see above).

References

Carson D.J., 1999. Climate Modelling: achievements and prospects, Quarterly Journal of the Royal Meteorological Society, **125**, 1-27.

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. and Jarvis A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978. URL: www.worldclim.org (date last accessed: 29 June 2009).

Houghton J.T., Ding Y., Griggs D.J., editors, 2001. Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

Jensen M.E. and Haise H.R., 1963. Estimation of evapotranspiration from solar radiation. *Journal of Irrigation and Drainage Division, Proceedings of the American Society of Civil Engineers*, **89**, 15-41.

Mitchell T.D., Carter T.R., Jones P.D., Hulme M., and New M., 2003. A comprehensive set of climate scenarios for Europe and the globe. *Tyndall Centre Working Paper*, **55**, 1-25. URL: http://www.cru.uea.ac.uk/cru/data/hrg.htm (date last accessed: 29 June 2009).

Mitchell T.D. and Jones P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. International Journal of Climatology, **25**, 693-712.

Piedallu C. and Gegout J.C., 2007. Multiscale computation of solar radiation for predictive vegetation modelling. *Annals of Forest Science*, **64** (8), 899-909.

United States Geological Survey (USGS), 2003. Seamless Shuttle Radar Topography Mission (SRTM) Finished 3 Arc Second (~90 meters).
URL: http://seamless.usgs.gov/website/seamless/products/srtm3arc.asp (date last accessed: 29 June 2009).

Zimmermann NE, Kienast F (1999) Predictive mapping of alpine grasslands in Switzerland: species versus community approach. Journal of Vegetation Science, 10, 469–482.

Supplementary material S2: weighted average ensemble forecasting.

Ensemble forecasting basically consists in combining projections of different models in order to overcome the variability arising from individual modeling techniques and provide more robust projections (Araújo and New 2007).

In this study, we employed five different modeling techniques: generalized linear models (GLM; McCullagh and Nelder 1989), generalized additive models (GAM; Hastie and Tibshirani 1986), generalized boosted models (GBM; Ridgeway 1999), random forest (RF; Breiman 2001) and multivariate adaptive regression splines (MARS; Friedman 1991, Friedman et al. 2000). For each species and in each of the study areas, the results yielded by these five modeling techniques were evaluated through a data-splitting procedure (see Thuiller et al. 2009 for detailed explanations): a model was trained on 70% of the data before being evaluated on the remaining 30%. This splitting was done while maintaining the original prevalence of a species' presences and absences in the data. The agreement between predicted and observed values was then evaluated using two different measures: the area under the relative operating characteristic curve (AUC; Hanley and McNeil 1982) and the True Skill Statistic (TSS; Allouche et al. 2006). This data-splitting procedure was repeated 50 times and the evaluation values averaged.

Note that, while model evaluation was done using the above-mentioned data-splitting procedure, the final models used for spatial projections were calibrated using 100% of the data for a species in a given study area, thereby allowing to take advantage of the full available data.

To obtain a "weighted average" projection, the individual projection yielded by each of the five modeling techniques (GLM, GAM, GBM, RF and MARS) were weighted by either their AUC or TSS evaluation score and averaged. The idea behind this weighting method, which was shown to be particularly robust (Marmion et al., 2009), is that models achieving higher evaluation scores should contribute more to the ensemble forecasting that those obtaining poorer evaluations.

Let $Model_i$ be the probabilistic projection obtained for a pixel through modeling technique i, AUC_i the AUC value obtained by the data splitting procedure for modeling technique i and TSS_i the TSS value obtained by the data splitting procedure for modeling technique i. The value of i belongs to the interval [1:5], representing each of the five individual modeling techniques (i.e. GLM, GAM, GBM, RF and MARS). For each species, two weighted average (WA) projections, based on respectively the AUC (WA_{AUC}) or the TSS (WA_{TSS}) values, were then computed as follows (Eq. S2.1 – S2.2):

$$WA_{AUC} = \frac{\sum_{i=1}^{5} Model_i \times (AUC_i - 0.5) \times 2}{\sum_{i=1}^{5} (AUC_i - 0.5) \times 2} \qquad \text{Eq. S2.1}$$

$$WA_{TSS} = \frac{\sum_{i=1}^{5} Model_i \times TSS_i}{\sum_{i=1}^{5} TSS_i} \qquad \text{Eq. S2.2}$$

Note that, to avoid basing projection on poorly calibrated models, models with an AUC value <= 0.7 or a TSS value <= 0.4 were disregarded in the summation of Eq. S2.1 and Eq. S2.2.

The WA_{AUC} and WA_{TSS} probabilistic projections in the range [0:1] were then reclassified into binary projection: 1 or 0 for respectively suitable or unsuitable habitat. This was done by computing a "weighted average" threshold (Eq. S2.3 and S2.4) and reclassifying projections greater or equal to the threshold as 1 (suitable habitat) and those below as 0 (unsuitable habitat).

$$Threshold_{WA\,AUC} = \frac{\sum_{i=1}^{5} Threshold_i \times (AUC_i - 0.5) \times 2}{\sum_{i=1}^{5} (AUC_i - 0.5) \times 2} \qquad \text{Eq. S2.3}$$

$$Threshold_{WA\,TSS} = \frac{\sum_{i=1}^{5} Threshold_i \times TSS_i}{\sum_{i=1}^{5} TSS_i} \qquad \text{Eq. S2.4}$$

References

Araújo M.B. and New M., 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42-47.

Allouche O., Tsoar A. and Kadmon R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223-1232.

Breiman L., 2001. Random Forests. *Machine Learning*, **45**, 5-32.

Friedman J., 1991. Multivariate adaptive regression splines. *The Annals of Statistics*, **19**, 1-141.

Ridgeway G., 1999. The state of boosting. *Computing Science and Statistics*, **31**, 172-181.

Friedman J., Hastie T. and Tibshirani R., 2000. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, **28**, 337-407.

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. and Tibshirani R., 1986. Generalized additive models. *Statistical Science*, **1**, 297-318.

McCullagh P. and Nelder J. A., 1989. Generalized linear models, 2nd ed. Chapman and Hall.

Marmion M., Parviainen M., Luoto M., Heikkinen R.K. and Thuiller W., 2008. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59-69.

Supplementary material S3: Vegetation belt definition and species classification

Definition of vegetation belt limits

To divide each study area into different altitudinal life zones, the four following vegetation belts were considered (Theurillat 1991, Körner 2003):

Alpine: Life zone encompassing exclusively vegetation above the upper limit of the natural tree-line. Only grasslands or small shrubs such as dwarf *Salix sp*. are found in this vegetation belt. The vegetation period lasts for ~50-100 days per year.

Subalpine: Life zone between the closed Montane forest and the uppermost limit of small tree species individuals. This zone represents the transition zone between fully grown forest and Alpine grasslands. Deciduous trees are mostly absent from this vegetation belt dominated by conifers. The vegetation period lasts for ~100-200 days per year.

Montane: Life zone where the native vegetation is mainly composed of fully grown coniferous forest, or mixed forests with deciduous trees such as *Fagus sp*. The vegetation period lasts for ~200-250 days per year.

Coline: Lowest and hence warmest life zone where the native vegetation is mainly deciduous forests composed of species such as *Quercus sp*., *Fraxinus sp*., or *Acer sp*. The vegetation period lasts for more than 250 days per year.

Although the upper and lower limits of vegetation belts are often defined in terms of elevation, essentially, a vegetation belt corresponds to a certain temperature range (Theurillat 1991). Since the different study areas considered in our analysis are scattered over a fairly large latitudinal and longitudinal gradient, a single set of vegetation belt limits based on elevation or even temperature cannot provide satisfactory classification.

We therefore defined individually the vegetation belt limits for each study area by looking at literature specific to a given mountain range (Table S3.1). When the limits found in the literature were given in terms of elevation, which was mostly the case, we converted these elevation values into mean annual temperatures. This was done by looking at the distribution of mean annual temperature of pixels having the threshold's elevation within the study area. Using the mean annual temperature thresholds we could then classify each pixel of a study area as belonging either to the Alpine, Subalpine, Montane or Coline vegetation belt. Using temperature rather than elevation for delimiting vegetation belts also offers the advantage of accounting for variable topographic position within a study area: typically, in the northern hemisphere, the limit of a given vegetation belt will be at lower elevation on north-facing than on south-facing slopes.

Table S3.1. Range of vegetation belts for each study area given in terms of elevation above sea level and mean annual temperature. A = Alpine, SA = Subalpine, M = Montane, C = Coline.

Study Area name	Vegetation belt thresholds by elevation [m a.s.l.]	Vegetation belt thresholds by mean annual temperature [°C]	Reference
Austrian Alps	**A:** > 1800 m **SA:** 1200 – 1800 m **M:** 500 – 1200 m **C:** < 500 m	**A:** < 1 °C **SA:** 1 – 4 °C **M:** 4 – 7.5 °C **C:** > 7.5 °C	Dullinger et al. (2003)
Carpathians	**A:** > 2200 m **SA:** 1700 – 2200 m **M:** 700 – 1700 m **C:** < 700 m	**A:** < 0 °C **SA:** 0 – 2 °C **M:** 2 – 7 °C **C:** > 7 °C	Coldea (1991)
French Alps 1 and 2	**A:** > 2400 m **SA:** 1600 – 2400 m **M:** 800 – 1600 m **C:** < 800 m	**A:** < 3 °C **SA:** 3 – 6 °C **M:** 6 – 10 °C **C:** > 10 °C	Rameau et al. (2003)
Italian Apennines	**A:** > 2300 m **SA:** 1600 – 2300 m **M:** 900 – 1600 m **C:** < 900 m	**A:** < 2.5 °C **SA:** 2.5 – 7.5 °C **M:** 7.5 – 11.5 °C **C:** > 11.5 °C	Stanisci (2008) Theurillat et al. (2009)
Norwegian Scandes	**A:** > 1000 m **SA:** 500 – 1000 m **M:** < 500 m **C:** NA	**A:** < 1 °C **SA:** 1 – 4 °C **M:** > 4 °C **C:** NA	Grytnes J-A. (pers. com)
Pyrenees 1 and 2	**A:** > 2300 m **SA:** 1800 – 2300 m **M:** 500 – 1800 m **C:** < 500 m	**A:** < 3 °C **SA:** 3 – 6 °C **M:** 6 – 12 °C **C:** > 12 °C	Rivas-Martínez (1987)
Scottish Highlands	**A:** > 600 m **SA:** < 600 m **M:** NA **C:** NA	**A:** < 5 °C **SA:** > 5 °C **M:** NA **C:** NA	Horsfield & Thompson (1996)
Swiss inner Alps 1 and 2	**A:** > 2400 m **SA:** 1600 – 2400 m **M:** 800 – 1600 m **C:** < 800 m	**A:** < 0 °C **SA:** 0 – 4 °C **M:** 4 – 8 °C **C:** > 8 °C	Aeschimann and Burdet (1994)
Swiss Western Alps	**A:** > 1800 m **SA:** 1300 – 1800 m **M:** 700 – 1300 m **C:** < 700 m	**A:** < 3 °C **SA:** 3 – 5.5 °C **M:** 5.5 – 8.5 °C **C:** > 8.5 °C	Aeschimann and Burdet (1994)

Classification of species into vegetation belts

For a given study area, a species was associated to the vegetation belt in which most of its occurrence records fell into.

References

Aeschimann D. and Burdet H., 1994. Flore de la Suisse et des territoires limitrophes, le nouveau Binz. 2nd edition, Editions du Griffon, Neuchâtel, Switzerland.

Coldea G., 1991. Prodrome des associations végétales des Carpates du Sud-Est (Carpates Roumaines). *Documents Phytosociologiques*, **13**, 317-539.

Dullinger S., Dirnböck T., Greimler J. and Grabherr G., 2003. A resampling approach for evaluating effects of pasture abandonment on subalpine plant species diversity. *Journal of Vegetation Science*, **14**, 243-252.

Horsfield D. and Thompson D.B.A., 1996. The Uplands: guidance on terminology regarding altitudinal zonation and related terms. Information and Advisory Note No. 26. SNH, Battleby.

Rameau J.C., Mansion D. and Dumé G., 1993. Flore forestière française - guide écologique illustré - tome 2: montagnes. Institut pour le Développement, Paris.

Rivas-Martínez, S. 1987. Memorias del mapa de vegetación de España. Ministerio de Agricultura, Pesca y Alimentación. Madrid.

Stanisci A., 2008. L'Apennino centrale. In: Gerdol R., Tomaselli M. and Stanisci A. La vegetazione delle montagne italiane, p. 299-334, Club Alpino Italiano.

Theurillat J.P., 1991. Les étages de végétation dans les Alpes centrales occidentales (Vegetation levels in the western Central Alps). *Saussurea*, **22**. 103-147.

Theurillat J.-P., Iocchi M., Cutini M. and De Marco G., 2009. Vascular plant richness along an elevation gradient at Monte Velino (Central Apennines, Italy). *Biogeografia*, 28, *in press*.

Supplementary material S4: Evaluation of species distribution models, comparison of evaluations across modeling methodologies and comparison of spatial projections across modeling methodologies.

Methods reminder

For each species in each dataset, we modeled potential distribution using five different modeling techniques: generalized linear models (GLM), generalized additive models (GAM), generalized boosted models (GBM), random forest (RF) and multivariate adaptive regression splines (MARS). Finally all projections were also combined in an ensemble forecasting approach to produce a "weighted average" (WA) model.

The predictive power of each model was evaluated through a data-splitting procedure (see Thuiller et al. 2009 for detailed explanations): a model was trained on 70% of the data before being evaluated on the remaining 30%. This splitting was done while maintaining the original prevalence of a species' presences and absences in the data. The agreement between predicted and observed values was then evaluated using two different measures: the area under the relative operating characteristic curve (AUC; Hanley and McNeil 1982) and the True Skill Statistic (TSS; Allouche et al. 2006). This data-splitting procedure was repeated 50 times and the evaluation values averaged (all values presented hereafter are these averages). The AUC evaluation value varies between 0.5 for a model whose prediction are no better than random, to 1 for a model achieving perfect agreement with the observed data. The TSS evaluation value varies between 0 (random model) and 1 (perfect agreement).

Finally, we reclassified the probabilistic projections of each model into binary values, representing either suitable or unsuitable habitat. This conversion required the selection of a threshold above which a pixel was reclassified as potentially suitable and unsuitable below. We tested two different thresholds: (i) maximizing jointly the percentage of presences and absences correctly predicted, i.e. the probability where sensitivity = specificity (Liu et al. 2005). This threshold is thereafter referred to as "AUC-based threshold". (ii) maximizing the value of the TSS evaluation value, thereafter referred to as "TSS-based threshold".

Evaluation of species distribution models predictive power using AUC and TSS measures

The results of this evaluation procedure (AUC and TSS values) are presented for each modeling technique and each study area in Figure S4.1, Table S4.1 and Table S4.2 (next pages). All presented values are averages over the 3095 modeled species.

With an Average AUC of 0.78 ±0.7 (1SD), GBM is the technique that earns the highest evaluation scores. It is followed by RF (average AUC: 0.77 ±0.7) and GLM (average AUC: 0.76 ±0.7). MARS (average AUC: 0.73 ±0.7) and GAM (average AUC: 0.70 ±0.8) are the least well performing methods among the five. The exact same trend is found when looking at TSS evaluation measures (Table S4.1).

Given that the difference in AUC and TSS evaluation values between the three highest ranked modeling techniques (GBM, RF and GLM) is minim, these should probably be considered as offering equal performance.

Looking at the proportion of useful models (a useful model being here defined as a model with AUC > 0.7 or TSS > 0.4), the observed trend in model ranking is similar: GBM produced the highest proportion of useful models, with 76.8 ±21.7% of the species having models evaluated above the AUC threshold and 58.9 ±27.4% above the TSS threshold. In 10 out of 12 study areas, GBM is the method that had the highest percentage of species being evaluated above either the AUC or the TSS threshold.

RF models and GLM generally ranked second or third, while MARS generally ranked fourth and GAM was the method getting the lowest percentages of models rising above the usefulness threshold.

Table S4. 1. Average AUC and TSS value of all species for each study area and each modeling technique. Mean and Standard deviation per modeling technique are also given in the two rightmost columns.

	Modeling technique	Austrian Alps	French Alps 1	French Alps 2	Swiss Inner Alps 2	Italian Apennines	Pyrenees 1	Swiss Western Alps	Norwegian Scandes	Pyrenees 2	Scottish Highlands	Carpathians	Swiss Inner Alps 1	**Mean**	**Standard Deviation**
Mean AUC value	**GLM**	0.74	0.86	0.66	0.77	0.74	0.63	0.82	0.85	0.77	0.77	0.74	0.74	0.76	0.07
	GAM	0.70	0.84	0.60	0.71	0.68	0.57	0.80	0.82	0.65	0.65	0.67	0.69	0.70	0.08
	GBM	0.76	0.90	0.70	0.77	0.75	0.65	0.82	0.88	0.79	0.79	0.75	0.78	0.78	0.07
	RF	0.75	0.91	0.70	0.75	0.74	0.64	0.81	0.87	0.80	0.80	0.75	0.72	0.77	0.07
	MARS	0.71	0.84	0.65	0.73	0.70	0.62	0.78	0.84	0.72	0.72	0.72	0.73	0.73	0.07
Mean TSS value	**GLM**	0.46	0.65	0.35	0.52	0.47	0.32	0.60	0.66	0.58	0.58	0.46	0.47	0.51	0.11
	GAM	0.38	0.62	0.19	0.41	0.35	0.15	0.57	0.59	0.32	0.32	0.33	0.36	0.38	0.15
	GBM	0.49	0.72	0.42	0.53	0.48	0.35	0.59	0.70	0.60	0.60	0.48	0.53	0.54	0.11
	RF	0.45	0.72	0.41	0.43	0.46	0.34	0.57	0.68	0.61	0.61	0.45	0.40	0.51	0.12
	MARS	0.42	0.62	0.34	0.43	0.42	0.28	0.54	0.65	0.50	0.50	0.43	0.45	0.47	0.11
Number of species/models		269	597	114	100	10	1118	287	90	5	124	116	265	3095	3095

Table S4. 2. Percentage of species with distribution model considered as useful (i.e., AUC > 0.7 and TSS > 0.4) in each study area and for each modeling technique. Mean and Standard deviation per modeling technique are also given in the two rightmost columns.

	Modeling technique	Austrian Alps	French Alps 1	French Alps 2	Swiss Inner Alps 2	Italian Apennines	Pyrenees 1	Swiss Western Alps	Norwegian Scandes	Pyrenees 2	Scottish Highlands	Carpathians	Swiss Inner Alps 1	**Mean**	**Standard Deviation**
% of species with AUC > 0.7	**GLM**	64.7	96.5	28.9	68.0	20.0	68.4	96.9	89.9	80	62.1	62.1	81.0	68.2	24.0
	GAM	55.4	93.6	17.5	48.0	10.0	54.8	90.6	82.0	40	43.5	48.3	65.4	54.1	26.1
	GBM	72.5	100	45.6	70.0	30.0	75.0	94.8	97.8	100	70.2	82.8	83.7	76.8	21.7
	RF	68.4	99.8	48.2	67.0	20.0	63.0	94.8	97.8	100	66.9	58.6	73.8	71.5	24.0
	MARS	58.7	97.0	31.6	48.0	10.0	59.4	88.9	94.4	80	54.0	62.1	76.4	63.4	26.0
% of species with TSS > 0.4	**GLM**	45.0	82.6	14.0	42.0	10.0	51.1	83.3	83.1	80	40.3	41.4	53.6	52.2	25.7
	GAM	37.2	78.7	9.6	26.0	0.0	43.6	74.6	74.2	40	28.2	31.0	40.7	40.3	24.9
	GBM	49.8	97.3	24.6	41.0	20.0	52.3	80.8	92.1	100	40.3	52.6	56.3	58.9	27.4
	RF	37.2	97.3	20.2	36.0	20.0	32.3	71.4	87.6	100	29.0	23.3	35.0	49.1	30.8
	MARS	29.7	81.1	12.3	28.0	10.0	36.4	64.5	83.1	60	31.5	37.1	38.8	42.7	24.3
Number of species/models		269	597	114	100	10	1118	287	90	5	124	116	265	3095	3095

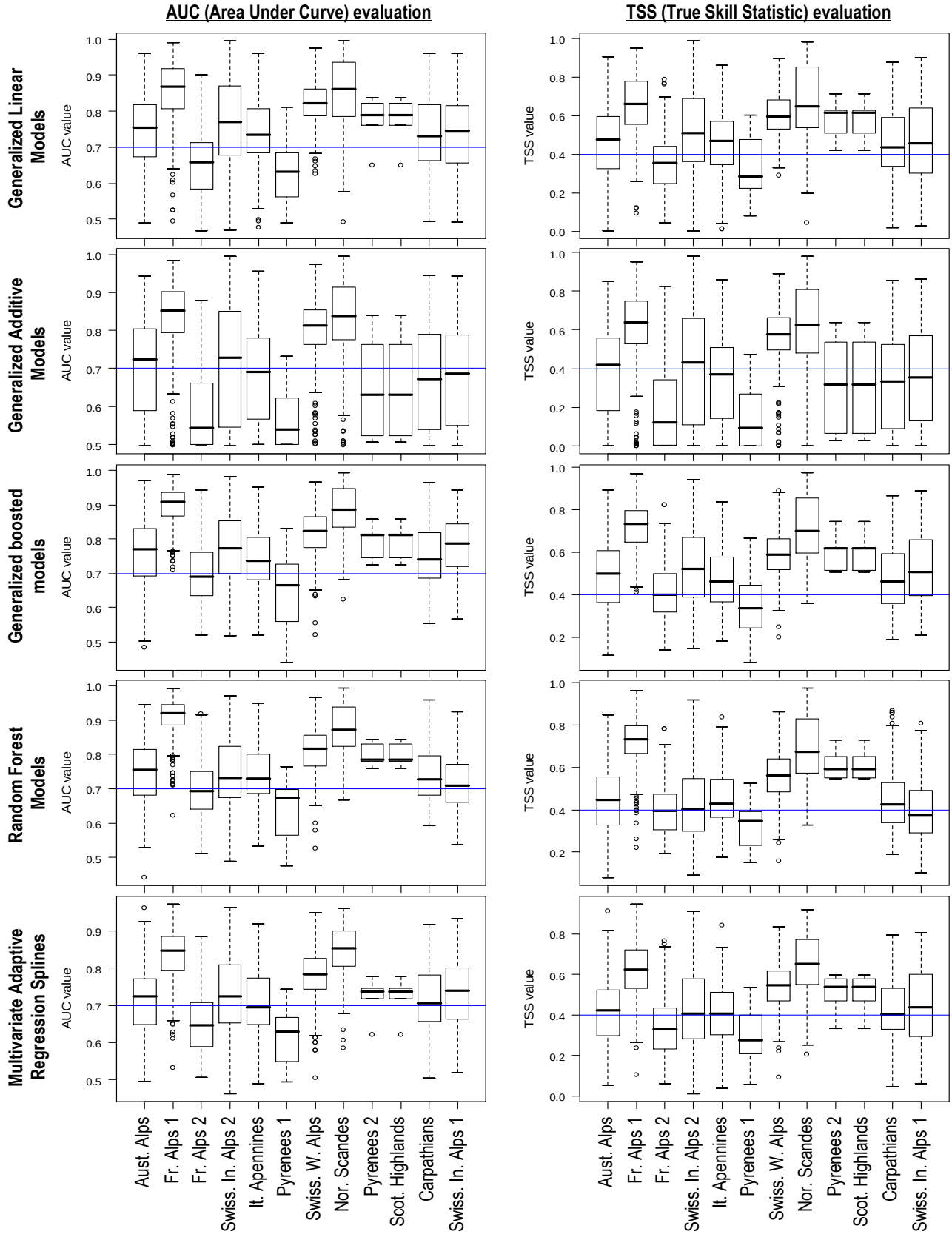


Figure S4.1. Boxplots of AUC and TSS evaluation values for each study area and each modeling technique. Blue line highlights the AUC = 0.7 and TSS = 0.4 thresholds above which a model was considered as useful and kept for analysis.

Comparison of AUC and TSS model evaluation measures

With a correlation varying between 95-98% (N=15'475) depending on the study area, the AUC and TSS measures provided quasi-equivalent model evaluation. In other words, models that obtained a high AUC evaluation almost always had a high TSS evaluation too. Indeed, across all modeling techniques, 94% ±1.9% (1 SD, N=15'475) of the time, these two evaluation measures agreed upon classifying a model as useful or not (i.e., AUC > 0.7 and TSS > 0.4).

Comparison of model evaluation across modeling techniques

Comparing evaluation values between modeling techniques reveals that in 78% of the cases (N=240), the correlation in evaluation measure (either AUC or TSS) between any two modeling techniques was > 80%. Furthermore, 91% ± 1.4% (1 SD, N=20) of the time, the evaluation measures between any two pairs of modeling technique agreed upon classifying a model as useful or not (i.e., AUC > 0.7 and TSS > 0.4). This means that species that were well modeled with one technique were generally also well modeled using another methodology, and those that had poor models with one technique were also poorly modeled when using another method. Such result is reassuring as it indicates that the underlying species biology or field data is more important in determining the quality of a distribution model than the particular technique that was used. Our result also corroborates with previous findings from Elith et al. (2006) in large methodological comparison study involving 226 species and 16 modeling methods.

Spatial agreement between projections obtained from AUC-based and TSS-based reclassification thresholds

For each modeling technique (GLM, GAM, GBM, RF, MARS and WA), we compared the binary spatial projections (i.e. habitat is suitable or not) obtained when using either the AUC-based or the TSS-based reclassification threshold. We term this comparison "spatial agreement" and it is the percentage of pixels, for a given species in a given study area, that have the same value, i.e. the compared pixels are either both suitable or both non-suitable. This spatial agreement was computed under current and the four future climate change scenarios (A1FI, A2, B1 and B2).

The spatial projections obtained using either the AUC-based or TSS-based reclassification threshold were highly similar: 94 ±11.3% (1 SD) of pixels were predicted identically across all study areas, climatic scenarios, species and modeling techniques (N=92'850). No important difference in this spatial agreement was observed across climatic scenarios or modeling techniques (Table S4.3). This means that projections of suitable habitat obtained by reclassifying the original probabilistic values with either the AUC-based or the TSS-based threshold yielded, in their large majority, very similar results.

Table S4.3. Mean and Standard deviation (SD) of the spatial agreement (% of identically predicted surface) between AUC-based and TSS-based reclassification thresholds. Mean and SD Values are based on 3095 × 6 models when computed by climate change and 3095 × 5 models when computed by modeling technique.

By Climatic Scenario			By modeling technique		
Climatic Scenario	Mean	SD	Modeling technique	Mean	SD
Present	93.7	9.4	GLM	96.6	4.9
A1FI	94.8	13.1	GAM	95.5	7.1
A2	94.8	11.8	GBM	96.6	7.8
B1	94.5	10.9	RF	91.5	16.4
B2	94.5	10.9	MARS	95.1	12.6
			WA	91.3	13.1

Spatial agreement between modeling techniques

To compare the spatial projections yielded by the different modeling techniques (GLM, GAM, GBM, RF, MARS and WA), we computed the spatial agreement as defined above between all modeling techniques. This spatial agreement was measured pair-wise for all species, all modeling techniques and all climatic scenarios (Present climate, A1FI, A2, B1 and B2).

GLM, GAM, GBM, RF and WA provided reasonably similar projections, with and average of ~75 ±25% (1 SD) spatial agreement between any two of these modeling techniques across all study areas and climatic scenarios (Table S4.4). MARS proved to be always the method with the least spatial agreement with any other method: on average, only ~55 ±32% (1 SD) of the pixels had the same value as projected by the method it was compared to.

Comparing the spatial agreement between modeling techniques across the different climatic conditions, showed that, on average, the agreement ~5.5 ±4.5% higher under current climatic conditions than under projected future climatic scenarios. More extreme climate change scenarios (e.g. A1FI) also showed lower levels of spatial agreement than less extreme ones (e.g. B1, Table S4.4), but this difference was rather small (~1.5 ±1.7%) and thus not of significance.

Interestingly, the WA method provided spatial projections having the highest spatial agreement with any other individual modeling technique (~85 ±21% similarity with GLM, GAM, GBM and RF and 60 ±33% with MARS). This result thus further supports the implementation of ensemble forecasting as an effective way of providing robust projections. It also provides strong support for the results from Marmion et al. (2009) who already found that WA is a robust consensus method.

Table S4.4. Mean ± standard error of the spatial agreement (% of identically predicted surface) between modeling techniques under **a)** all climatic conditions (average of present, A1FI, A2, B1 and B2), **b)** current climatic conditions, **c)** A1FI climate change scenario and, **d)** B1 climate change scenario.

a) All climatic conditions

	GLM	GAM	GBM	RF	MARS	WA
GLM	-	-	-	-	-	-
GAM	78.4 ± 26.5	-	-	-	-	-
GBM	77.4 ± 25.3	77.8 ± 25.7	-	-	-	-
RF	72.1 ± 26.7	71.3 ± 28.8	80.1 ± 25.4	-	-	-
MARS	56.9 ± 31.4	56.6 ± 32.4	55.2 ± 34.5	55.6 ± 33.8	-	-
WA	85.2 ± 19.5	81.6 ± 24.4	86.7 ± 19.3	82.8 ± 22.7	60.4 ± 33	-

b) Present climatic conditions

	GLM	GAM	GBM	RF	MARS	WA
GLM	-	-	-	-	-	-
GAM	83.4 ± 17.8	-	-	-	-	-
GBM	83 ± 12.3	78.7 ± 19.3	-	-	-	-
RF	75.6 ± 14.9	71.8 ± 21.9	85.8 ± 11.8	-	-	-
MARS	66.5 ± 24.8	64.6 ± 26.4	65.8 ± 27.7	62.5 ± 28.6	-	-
WA	85.4 ± 12.7	80 ± 21.3	90 ± 9.6	88.1 ± 12	68.2 ± 26.4	-

c) A1FI climatic change projections

	GLM	GAM	GBM	RF	MARS	WA
GLM	-	-	-	-	-	-
GAM	75.9 +- 31.1	-	-	-	-	-
GBM	74.4 +- 31.6	77.9 +- 29.8	-	-	-	-
RF	68.9 +- 33.2	70.1 +- 33.9	75.9 +- 31.9	-	-	-
MARS	53.9 +- 34.3	53.5 +- 35.4	51.6 +- 36.7	53.8 +- 35.6	-	-
WA	84.8 +- 23.9	82.6 +- 26.6	85 +- 24	79 +- 28.5	58 +- 35.4	-

d) B1 climatic change projections

	GLM	GAM	GBM	RF	MARS	WA
GLM	-	-	-	-	-	-
GAM	78 +- 25.9	-	-	-	-	-
GBM	77.1 +- 24.3	77.3 +- 25.3	-	-	-	-
RF	72.6 +- 25.6	71.9 +- 27.7	80.6 +- 24.3	-	-	-
MARS	54.5 +- 31.2	54.9 +- 32.1	52.6 +- 34.7	53.1 +- 34.4	-	-
WA	85.4 +- 18.6	81.6 +- 24.1	86.4 +- 18.8	83.1 +- 21.4	58.3 +- 33.5	-

References

Allouche O., Tsoar A. and Kadmon R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223-1232.

Elith J., Graham C.H., Anderson R.P., Dudik M., Ferrier S., Guisan A., Hijmans R. J., Huettmann F., Leathwick J.R., Lehmann A., Li J., Lohmann L.G., Loiselle B.A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J.M. and Peterson A.T., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29** (2), 129-151.

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.

Liu C.R., 2005. Selecting threshold of occurence in the prediction of species distributions. *Ecography*, **28**, 385-393.

Marmion M., Parviainen M., Luoto M., Heikkinen R.K. and Thuiller W., 2008. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59-69.

Thuiller W., Lafourcade B., Engler R. and Araújo M.B., 2009. BIOMOD – A platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369-373.

Supplementary material S5: Percentage of species, for each study area, with a projected decrease in distribution by 2100 of respectively 100%, 90% or 80% under A1FI, A2, B1 and B2 climate change scenarios.

The "mean" column represents the average value obtained from the AUC and TSS weighted average projections. The "80% range" column indicates the range of values observed in 80% of the individual models that yielded results closest to the average value. This is equivalent to showing the range of results yielded by all but the two most outlier models.

Table S5.1. Percentages of species with projected decrease in distribution by 2100 of 100%, 90% or 80% under A1FI climate change scenario (most extreme).

			100% range loss		>90% range loss		>80% range loss	
DataSet	TotalSp	RetainedSp	Mean	80% Range	Mean	80% Range	Mean	80% Range
Austrian Alps	269	200	54.8	21.9 - 65.1	61.6	36.9 - 74.9	62.6	37.5 - 75.9
French Alps 1	597	597	5.1	0.2 - 7.9	22.8	10.3 - 34.8	31.1	14.7 - 42.7
French Alps 2	114	63	40.5	22.2 - 71.2	47.6	29.6 - 75	47.6	29.6 - 75
Swiss Western Alps	287	281	27.8	2.1 - 41.5	44.1	12.5 - 60	47.3	16.7 - 62.9
Swiss inner Alps 1	265	231	33.3	0 - 50	33.3	0 - 66.7	50.0	0 - 100
Swiss inner Alps 2	100	79	31.8	12.3 - 45.1	70.9	45.5 - 93.3	75.3	57.1 - 93.8
Norwegian Scandes	116	88	16.6	0.4 - 39.1	37.2	7.8 - 64.9	45.9	22 - 68.4
Scottish Highlands	124	94	8.0	0 - 15.9	33.2	6.9 - 37.9	34.9	10.3 - 44.8
Carpathians	116	97	20.0	0 - 25	60.0	25 - 80	70.0	60 - 100
Pyrenees 1	1118	894	29.8	22.7 - 37	55.5	31.8 - 67.8	58.9	36.5 - 69
Pyrenees 2	5	5	29.5	1.5 - 53.5	41.8	13.9 - 65.6	44.6	13.9 - 65.6
Apennines	10	3	7.4	0.5 - 12.3	24.2	6.5 - 42.7	28.9	10 - 52.3

Table S5.2. Percentages of species with projected decrease in distribution by 2100 of 100%, 90% or 80% under B1 climate change scenario (least extreme).

			100% range loss		>90% range loss		>80% range loss	
DataSet	TotalSp	RetainedSp	Mean	80% Range	Mean	80% Range	Mean	80% Range
Austrian Alps	269	200	31.9	15.2 - 38	53.8	24.4 - 62.4	59.3	32.5 - 66.4
French Alps 1	597	597	0.1	0 - 0.5	3.3	0.6 - 5	6.1	1.3 - 10.4
French Alps 2	114	63	27.0	2.8 - 50	38.1	7.7 - 68.8	42.9	24.1 - 75
Swiss Western Alps	287	281	5.8	1.7 - 7.7	19.4	2.1 - 32.3	23.3	5.9 - 33.3
Swiss inner Alps 1	265	231	0.0	0 - 0	33.3	0 - 50	33.3	0 - 50
Swiss inner Alps 2	100	79	8.2	0.2 - 15.6	32.4	17.5 - 43.2	47.6	31.6 - 63
Norwegian Scandes	116	88	2.5	0.4 - 4.4	13.4	1.5 - 29.4	22.2	4.5 - 44.9
Scottish Highlands	124	94	1.2	0 - 2.7	9.7	1.2 - 17.5	12.0	2.3 - 22.5
Carpathians	116	97	0.0	0 - 0	20.0	0 - 20	40.0	20 - 80
Pyrenees 1	1118	894	11.2	9.1 - 18.4	27.7	23.4 - 33.3	31.5	27.3 - 36.5
Pyrenees 2	5	5	20.3	0 - 23.7	33.3	13.2 - 55.2	36.0	15.3 - 57.3
Apennines	10	3	4.9	0.5 - 9.9	11.0	2.6 - 17.3	15.7	0.5 - 27.9

Table S5.3. Percentages of species with projected decrease in distribution by 2100 of 100%, 90% or 80% under A2 climate change scenario (intermediate).

			100% range loss		>90% range loss		>80% range loss	
DataSet	TotalSp	RetainedSp	Mean	80% Range	Mean	80% Range	Mean	80% Range
Austrian Alps	269	200	46.9	17.1 - 60.4	60.1	31.3 - 65.8	62.9	34.4 - 67.2
French Alps 1	597	597	1.3	0 - 1.8	11.8	4.4 - 20.8	19.5	9.9 - 27.9
French Alps 2	114	63	34.1	5.6 - 68.8	42.1	12.8 - 71.2	45.2	25.9 - 75
Swiss Western Alps	287	281	14.8	0 - 23.1	36.9	7.8 - 54.3	41.5	8.6 - 55.7
Swiss inner Alps 1	265	231	16.7	0 - 33.3	33.3	0 - 50	33.3	0 - 50
Swiss inner Alps 2	100	79	18.4	3.8 - 31.2	55.7	38.6 - 67.3	67.5	46.5 - 79.5
Norwegian Scandes	116	88	11.3	0 - 22.1	32.4	11.2 - 57	39.5	16.1 - 62.7
Scottish Highlands	124	94	8.0	0 - 15.9	29.2	8.1 - 35.6	33.2	9.2 - 42.5
Carpathians	116	97	0.0	0 - 0	40.0	0 - 100	50.0	20 - 100
Pyrenees 1	1118	894	29.8	22.4 - 37	55.2	33.7 - 66.7	58.1	36.1 - 67.8
Pyrenees 2	5	5	31.1	11.1 - 53.1	37.6	12.5 - 59.4	41.4	13.9 - 62.5
Apennines	10	3	5.4	0.5 - 10.3	13.9	0 - 17.2	22.6	0.5 - 40.2

Table S5.4. Percentages of species with projected decrease in distribution by 2100 of 100%, 90% or 80% under B2 climate change scenario (intermediate but close to B1).

			100% range loss		>90% range loss		>80% range loss	
DataSet	TotalSp	RetainedSp	Mean	80% Range	Mean	80% Range	Mean	80% Range
Austrian Alps	269	200	35.2	17.7 - 38.5	53.2	27.5 - 61.7	59.1	36.3 - 65.8
French Alps 1	597	597	0.3	0 - 0.5	3.6	0.7 - 5.9	8.3	3.9 - 12.2
French Alps 2	114	63	22.2	3.7 - 56.3	35.7	10.3 - 62.5	40.5	11.1 - 65.4
Swiss Western Alps	287	281	5.8	1.7 - 10	21.4	6.9 - 38.6	27.3	6.3 - 44.6
Swiss inner Alps 1	265	231	0.0	0 - 0	16.7	0 - 33.3	16.7	0 - 33.3
Swiss inner Alps 2	100	79	9.5	0.2 - 18	31.9	19.9 - 41.5	47.5	36.1 - 57.6
Norwegian Scandes	116	88	5.0	0 - 9.9	17.9	2.2 - 39.1	26.5	6.7 - 50
Scottish Highlands	124	94	1.2	0 - 7.3	12.6	2.3 - 21.3	19.4	5.8 - 27.5
Carpathians	116	97	0.0	0 - 0	30.0	0 - 50	40.0	20 - 80
Pyrenees 1	1118	894	10.7	9.1 - 17.2	27.0	20.6 - 34.9	32.7	27.7 - 36.5
Pyrenees 2	5	5	19.8	0 - 21.7	34.4	12.5 - 57.3	37.0	12.5 - 60.4
Apennines	10	3	4.2	0 - 5	10.3	3.2 - 15.9	13.0	0 - 23.5

Supplementary material S6: Percentage of species, for each vegetation band, with a projected decrease in distribution by 2100 of respectively 100%, 90% or 80% under A1FI, A2, B1 and B2 climate change scenarios.

The "mean" column represents the average value obtained from the AUC and TSS weighted average projections. The "80% range" column indicates the range of values observed in 80% of the individual models that yielded results closest to the average value, This is equivalent to showing the range of results yielded by all but the two most outlier models.

Table S6.1. Percentages of species with projected decrease in distribution by 2100 of 100%, 90% or 80% under A1FI climate change scenario (most extreme).

				100% range loss		>90% range loss		>80% reange loss	
DataSet	VegBelt	TotalSp	RetainedSp	Mean	80% Range	Mean	80% Range	Mean	80% Range
Austrian Alps	Alpine	47	47	89.4	65 - 100	92.6	73.8 - 100	92.6	74.4 - 100
	SubAlpine	222	153	43.9	12.7 - 56.7	51.9	23.7 - 66.9	53.2	24.6 - 68.2
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
French Alps 1	Alpine	0	0	-	-	-	-	-	-
	SubAlpine	38	38	40.8	13.2 - 55.3	90.8	68.4 - 100	98.7	79 - 100
	Montane	224	224	5.6	0 - 11.2	35.1	11 - 51.8	49.1	20.5 - 59.4
	Coline	335	335	0.8	0 - 1.3	6.9	1 - 12	11.3	2.7 - 14.9
French Alps 2	Alpine	108	57	43.0	25 - 76.1	50.9	33.3 - 80.4	50.9	33.3 - 80.4
	SubAlpine	6	6	16.7	0 - 33.3	16.7	0 - 33.3	16.7	0 - 33.3
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Swiss Western Alps	Alpine	127	126	24.3	0.9 - 52.5	45.9	11.2 - 70.8	57.8	35.9 - 92.5
	SubAlpine	80	75	14.8	0 - 38.4	48.4	12.7 - 79.2	56.4	21.1 - 80.6
	Montane	74	74	6.8	0 - 13.5	14.2	1.4 - 24.3	18.9	2.7 - 28.4
	Coline	6	6	0.0	0 - 0	0.0	0 - 0	0.0	0 - 16.7
Swiss inner Alps 1	Alpine	161	130	5.8	0 - 6.7	31.7	6.4 - 54.2	39.5	10.1 - 69.2
	SubAlpine	102	101	9.4	0 - 15	15.3	5.6 - 29	16.3	1.2 - 24
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Swiss inner Alps 2	Alpine	34	28	44.6	5.6 - 75	64.3	19.1 - 88	67.9	22.2 - 92
	SubAlpine	66	51	18.2	0 - 29.3	32.6	11.4 - 44.4	35.6	13.3 - 50
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Norwegian Scandes	Alpine	74	73	9.7	0 - 19.4	40.0	8.3 - 45.8	42.1	12.5 - 52.8
	SubAlpine	15	15	0.0	0 - 0	0.0	0 - 0	0.0	0 - 6.7
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Scottish Highlands	Alpine	93	63	40.5	29.6 - 51.6	59.9	48.7 - 73.1	63.3	52.7 - 74.6
	SubAlpine	31	31	10.0	0 - 17.9	47.6	0 - 57.1	50.9	4.4 - 60
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Carpathians	Alpine	11	11	72.7	20 - 90.9	72.7	30 - 90.9	72.7	40 - 90.9
	SubAlpine	104	85	24.0	0 - 25.9	38.1	3.6 - 63.1	41.3	3.6 - 63.1
	Montane	1	1	0.0	0 - 0	0.0	0 - 0	0.0	0 - 0
	Coline	0	0	-	-	-	-	-	-
Pyrenees 1	Alpine	4	4	100.0	75 - 100	100.0	100 - 100	100.0	100 - 100
	SubAlpine	168	168	79.8	36.3 - 99.4	98.8	78.6 - 100	99.4	83.8 - 100
	Montane	935	711	20.3	2.3 - 33.6	65.0	33.7 - 91.8	70.4	46 - 92.4
	Coline	11	11	0.0	0 - 9.1	0.0	0 - 9.1	0.0	0 - 9.1
Pyrenees 2	Alpine	0	0	-	-	-	-	-	-
	SubAlpine	4	4	25.0	0 - 50	75.0	33.3 - 100	87.5	50 - 100
	Montane	1	1	0.0	0 - 0	0.0	0 - 100	0.0	0 - 100
	Coline	0	0	-	-	-	-	-	-
Apennines	Alpine	9	2	50.0	0 - 100	50.0	0 - 100	75.0	0 - 100
	SubAlpine	1	1	0.0	0 - 0	0.0	0 - 100	0.0	0 - 100
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-

Table S6.2. Percentages of species with projected decrease in distribution by 2100 of 100%, 90% or 80% under B1 climate change scenario (least extreme).

				100% range loss		>90% range loss		>80% reange loss	
DataSet	VegBelt	TotalSp	RetainedSp	Mean	80% Range	Mean	80% Range	Mean	80% Range
Austrian Alps	Alpine	47	47	85.1	50 - 93.6	94.7	66.7 - 100	94.7	71.4 - 100
	SubAlpine	222	153	15.3	0.9 - 20.3	40.9	9.3 - 52.9	48.3	18.6 - 58.7
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
French Alps 1	Alpine	0	0	-	-	-	-	-	-
	SubAlpine	38	38	0.0	0 - 7.9	22.4	2.6 - 42.1	34.2	7.9 - 60.5
	Montane	224	224	0.0	0 - 0	3.4	0 - 5.4	8.3	0 - 14.3
	Coline	335	335	0.2	0 - 0.3	1.0	0.3 - 0.6	1.5	0.3 - 2.3
French Alps 2	Alpine	108	57	29.8	4.2 - 55.6	41.2	18.8 - 72.2	45.6	27.1 - 78.3
	SubAlpine	6	6	0.0	0 - 0	8.3	0 - 33.3	16.7	0 - 33.3
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Swiss Western Alps	Alpine	127	126	2.4	0 - 4.1	18.4	1.7 - 41.5	34.7	8.6 - 50
	SubAlpine	80	75	4.0	0 - 5.6	14.8	1.4 - 30.1	18.2	1.4 - 45.2
	Montane	74	74	1.4	0 - 2.7	4.7	1.4 - 10.8	6.8	1.4 - 12.2
	Coline	6	6	0.0	0 - 0	0.0	0 - 16.7	0.0	0 - 16.7
Swiss inner Alps 1	Alpine	161	130	3.7	0 - 4.4	11.0	1.5 - 14.2	19.8	8.7 - 35.8
	SubAlpine	102	101	6.4	0 - 9	10.9	3.4 - 21	10.9	1.2 - 20
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Swiss inner Alps 2	Alpine	34	28	10.7	0 - 14.3	39.3	4.4 - 68.8	44.6	8.7 - 80
	SubAlpine	66	51	3.0	0 - 4.7	8.1	0 - 11.6	11.1	5.4 - 14
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Norwegian Scandes	Alpine	74	73	1.4	0 - 3.5	11.0	1.5 - 20.7	13.8	1.5 - 20.7
	SubAlpine	15	15	0.0	0 - 0	3.3	0 - 6.7	3.3	0 - 6.7
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Scottish Highlands	Alpine	93	63	16.3	12.2 - 25.4	39.0	32.7 - 49.2	44.1	36.7 - 54.2
	SubAlpine	31	31	1.7	0 - 3.6	6.6	0 - 10.7	8.3	3.2 - 13.3
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Carpathians	Alpine	11	11	63.6	20 - 81.8	72.7	20 - 90.9	72.7	40 - 90.9
	SubAlpine	104	85	14.7	0 - 17.7	28.3	5.3 - 51.2	31.5	7.3 - 53.6
	Montane	1	1	0.0	0 - 0	0.0	0 - 0	0.0	0 - 0
	Coline	0	0	-	-	-	-	-	-
Pyrenees 1	Alpine	4	4	50.0	0 - 100	75.0	50 - 100	100.0	75 - 100
	SubAlpine	168	168	22.0	0.6 - 40.5	67.6	31.7 - 98.2	85.1	53 - 99.4
	Montane	935	711	4.9	0 - 9.6	24.1	12.6 - 32.2	38.7	23.8 - 54.3
	Coline	11	11	0.0	0 - 0	4.5	0 - 9.1	9.1	0 - 11.1
Pyrenees 2	Alpine	0	0	-	-	-	-	-	-
	SubAlpine	4	4	0.0	0 - 0	25.0	0 - 25	50.0	25 - 75
	Montane	1	1	0.0	0 - 0	0.0	0 - 0	0.0	0 - 100
	Coline	0	0	-	-	-	-	-	-
Apennines	Alpine	9	2	0.0	0 - 0	50.0	0 - 100	50.0	0 - 100
	SubAlpine	1	1	0.0	0 - 0	0.0	0 - 0	0.0	0 - 0
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-

Table S6.3. Percentages of species with projected decrease in distribution by 2100 of 100%, 90% or 80% under A2 climate change scenario (intermediate).

				100% range loss		>90% range loss		>80% reange loss	
DataSet	VegBelt	TotalSp	RetainedSp	Mean	80% Range	Mean	80% Range	Mean	80% Range
Austrian Alps	Alpine	47	47	85.1	60 - 93.6	93.6	69.2 - 100	93.6	71.8 - 100
	SubAlpine	222	153	34.9	4.2 - 51.9	49.6	19.5 - 57.7	53.3	21.2 - 59.6
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
French Alps 1	Alpine	0	0	-	-	-	-	-	-
	SubAlpine	38	38	10.5	0 - 21.1	73.7	47.4 - 92.1	93.4	71.1 - 100
	Montane	224	224	1.3	0 - 2.2	15.4	2.7 - 28.1	29.9	10.5 - 44.6
	Coline	335	335	0.2	0 - 0.3	2.4	0.9 - 3.3	4.2	1.2 - 6.6
French Alps 2	Alpine	108	57	36.0	6.1 - 66.7	44.7	29.2 - 76.1	48.2	29.2 - 78.3
	SubAlpine	6	6	16.7	0 - 33.3	16.7	0 - 33.3	16.7	0 - 33.3
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Swiss Western Alps	Alpine	127	126	11.6	0 - 26.8	42.3	18.1 - 67.5	53.4	22.4 - 70
	SubAlpine	80	75	17.5	0 - 30.1	39.7	10.8 - 65.8	46.3	13.9 - 75
	Montane	74	74	5.4	0 - 10.8	10.8	1.4 - 17.6	12.2	4.2 - 21.6
	Coline	6	6	0.0	0 - 0	0.0	0 - 0	0.0	0 - 16.7
Swiss inner Alps 1	Alpine	161	130	2.9	0 - 6.7	15.9	0 - 22.3	30.0	0 - 57.7
	SubAlpine	102	101	8.4	0 - 12	11.4	3.3 - 21	13.9	1.2 - 22
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Swiss inner Alps 2	Alpine	34	28	33.9	4.4 - 56	57.1	11.1 - 84	64.3	14.3 - 88
	SubAlpine	66	51	4.0	0 - 7	25.3	6.3 - 37.8	28.5	6.7 - 39.5
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Norwegian Scandes	Alpine	74	73	9.7	0 - 13.9	35.2	9.7 - 43.1	40.0	11.1 - 51.4
	SubAlpine	15	15	0.0	0 - 0	0.0	0 - 0	0.0	0 - 0
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Scottish Highlands	Alpine	93	63	44.7	35.9 - 54.2	62.7	50 - 72.9	67.2	51.3 - 74.6
	SubAlpine	31	31	1.7	0 - 10.3	41.0	4.4 - 53.6	41.0	4.4 - 53.6
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Carpathians	Alpine	11	11	72.7	20 - 90.9	77.3	20 - 90.9	77.3	40 - 90.9
	SubAlpine	104	85	25.8	1.8 - 48.8	32.6	3.6 - 56	37.0	7.3 - 59.5
	Montane	1	1	0.0	0 - 0	0.0	0 - 0	0.0	0 - 0
	Coline	0	0	-	-	-	-	-	-
Pyrenees 1	Alpine	4	4	100.0	50 - 100	100.0	100 - 100	100.0	100 - 100
	SubAlpine	168	168	56.0	13.7 - 86.9	98.5	78.6 - 100	99.4	85.1 - 100
	Montane	935	711	9.2	0 - 18.1	45.8	25.1 - 59.5	60.6	32.1 - 75
	Coline	11	11	0.0	0 - 0	0.0	0 - 9.1	0.0	0 - 9.1
Pyrenees 2	Alpine	0	0	-	-	-	-	-	-
	SubAlpine	4	4	0.0	0 - 0	50.0	25 - 100	62.5	33.3 - 100
	Montane	1	1	0.0	0 - 0	0.0	0 - 100	0.0	0 - 100
	Coline	0	0	-	-	-	-	-	-
Apennines	Alpine	9	2	25.0	0 - 50	50.0	0 - 100	50.0	0 - 100
	SubAlpine	1	1	0.0	0 - 0	0.0	0 - 0	0.0	0 - 0
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-

Table S6.4. Percentages of species with projected decrease in distribution by 2100 of 100%, 90% or 80% under B2 climate change scenario (intermediate but close to B1).

				100% range loss		>90% range loss		>80% reange loss	
DataSet	VegBelt	TotalSp	RetainedSp	Mean	80% Range	Mean	80% Range	Mean	80% Range
Austrian Alps	Alpine	47	47	88.3	60 - 93.6	92.6	71.8 - 100	95.8	78.6 - 100
	SubAlpine	222	153	18.6	1.7 - 22.2	40.9	13.6 - 51.9	47.6	21.2 - 57.7
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
French Alps 1	Alpine	0	0	-	-	-	-	-	-
	SubAlpine	38	38	2.6	0 - 7.9	29.0	5.3 - 47.4	54.0	18.4 - 76.3
	Montane	224	224	0.0	0 - 0	4.0	0 - 7.6	10.5	0.9 - 17.9
	Coline	335	335	0.2	0 - 0.3	0.5	0 - 1.7	1.7	0.3 - 3.6
French Alps 2	Alpine	108	57	24.6	4.2 - 50	39.5	12.1 - 64.7	44.7	16.7 - 69.6
	SubAlpine	6	6	0.0	0 - 33.3	0.0	0 - 33.3	0.0	0 - 33.3
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Swiss Western Alps	Alpine	127	126	6.4	0 - 11.2	23.9	3.4 - 50.8	39.5	11.1 - 58.3
	SubAlpine	80	75	6.1	0 - 9.5	22.2	2.8 - 38.4	26.2	5.6 - 52.1
	Montane	74	74	2.0	0 - 4.1	4.7	0 - 10.8	6.8	1.4 - 12.2
	Coline	6	6	0.0	0 - 0	0.0	0 - 16.7	0.0	0 - 16.7
Swiss inner Alps 1	Alpine	161	130	2.9	0 - 3.5	9.4	2.9 - 15	13.9	0 - 22.3
	SubAlpine	102	101	5.9	0 - 9	11.4	3.3 - 17	11.9	3.4 - 21
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Swiss inner Alps 2	Alpine	34	28	10.7	0 - 14.3	41.1	4.4 - 68.8	50.0	13 - 84
	SubAlpine	66	51	3.0	0 - 5	10.2	0 - 14	14.3	5.4 - 16.3
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Norwegian Scandes	Alpine	74	73	1.4	0 - 9	15.2	2.8 - 26.2	23.5	6.9 - 32.3
	SubAlpine	15	15	0.0	0 - 0	0.0	0 - 0	0.0	0 - 0
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Scottish Highlands	Alpine	93	63	16.3	12.2 - 25.4	40.5	30.2 - 52.5	45.8	38.8 - 56.4
	SubAlpine	31	31	0.0	0 - 3.3	1.7	0 - 6.7	8.3	3.2 - 14.3
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-
Carpathians	Alpine	11	11	59.1	20 - 90.9	72.7	20 - 90.9	72.7	40 - 90.9
	SubAlpine	104	85	14.7	0 - 16.5	29.6	3.6 - 53.6	32.6	5.5 - 57.1
	Montane	1	1	0.0	0 - 0	0.0	0 - 0	0.0	0 - 0
	Coline	0	0	-	-	-	-	-	-
Pyrenees 1	Alpine	4	4	75.0	0 - 100	100.0	75 - 100	100.0	100 - 100
	SubAlpine	168	168	26.2	0.6 - 48.8	81.5	45.5 - 99.4	94.3	68.5 - 99.4
	Montane	935	711	5.3	0 - 10.6	19.9	9.5 - 30.4	36.4	23.5 - 47.5
	Coline	11	11	0.0	0 - 0	0.0	0 - 9.1	0.0	0 - 9.1
Pyrenees 2	Alpine	0	0	-	-	-	-	-	-
	SubAlpine	4	4	0.0	0 - 0	37.5	0 - 50	50.0	25 - 75
	Montane	1	1	0.0	0 - 0	0.0	0 - 0	0.0	0 - 100
	Coline	0	0	-	-	-	-	-	-
Apennines	Alpine	9	2	0.0	0 - 0	25.0	0 - 50	25.0	0 - 50
	SubAlpine	1	1	0.0	0 - 100	0.0	0 - 100	0.0	0 - 100
	Montane	0	0	-	-	-	-	-	-
	Coline	0	0	-	-	-	-	-	-

Supplementary material S7: Projected species range change (SRC) for the year 2100 under A1FI and B1 climate change scenarios, by endemism status and vegetative growth form.

SRC represents the percentage of potential spatial distribution that a species is projected to lose (negative values) or win (positive values) as compared to its potential distribution under current climatic conditions. For instance, a value of "-100" indicates that a species is projected to lose 100% of its potential distribution, i.e. its distribution is projected to decrease by 100% as compared to the surface of its potential distribution under current climatic conditions. A value of "0" indicates that a species is projected to have the same surface of potential distribution as under current climate (note that its distribution could shift while keeping the same area). Finally a value of "100" or "200" indicates that a species is projected to respectively double (100% increase) or triple (200% increase) its potential distribution under the projected future climatic conditions.

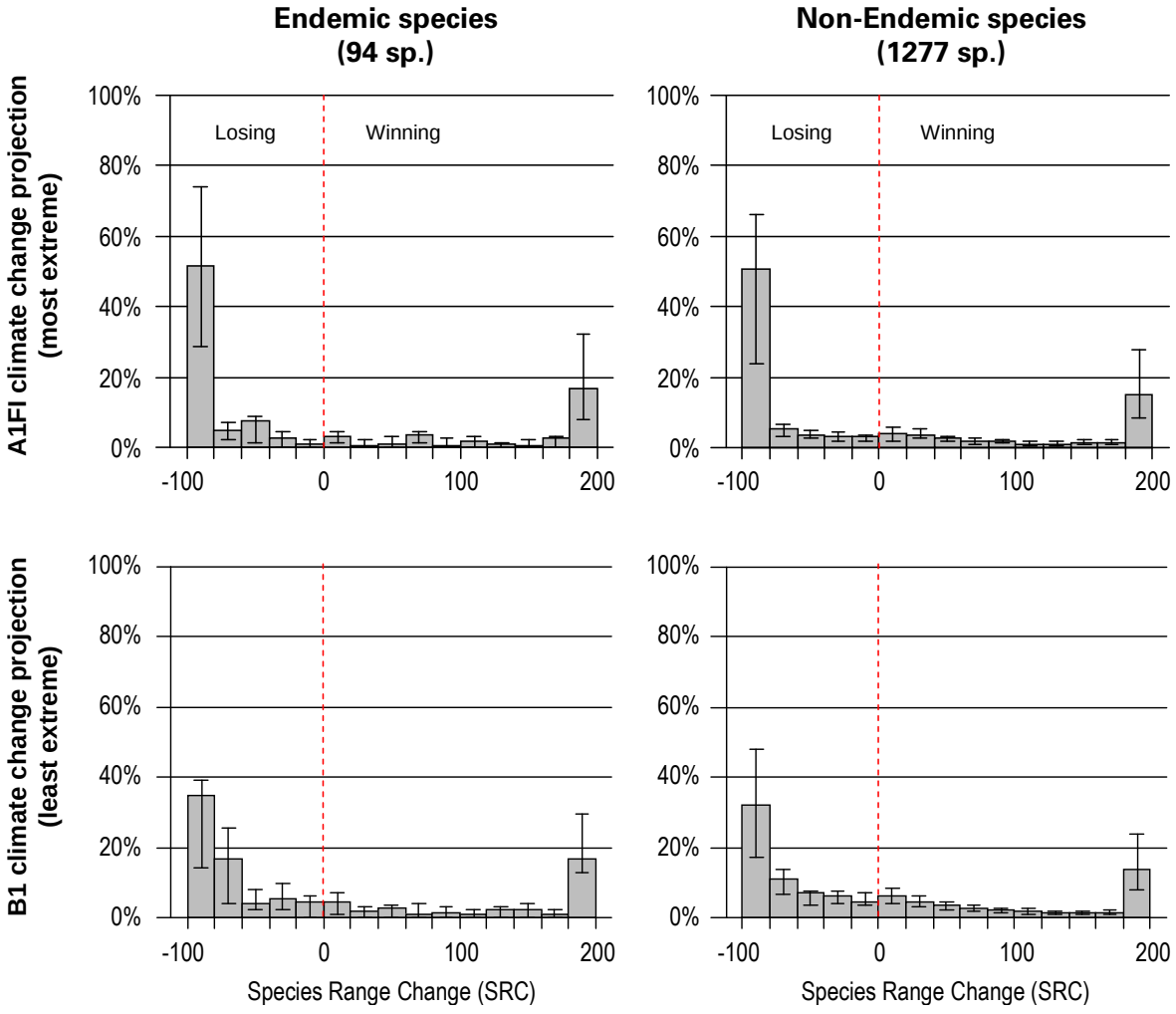


Figure S7.1. Projected SRC for the year 2100 under A1FI (upper panels) and B1 (lower panels) climate change scenarios for respectively endemic (left panels) and non-endemic species (right panels).

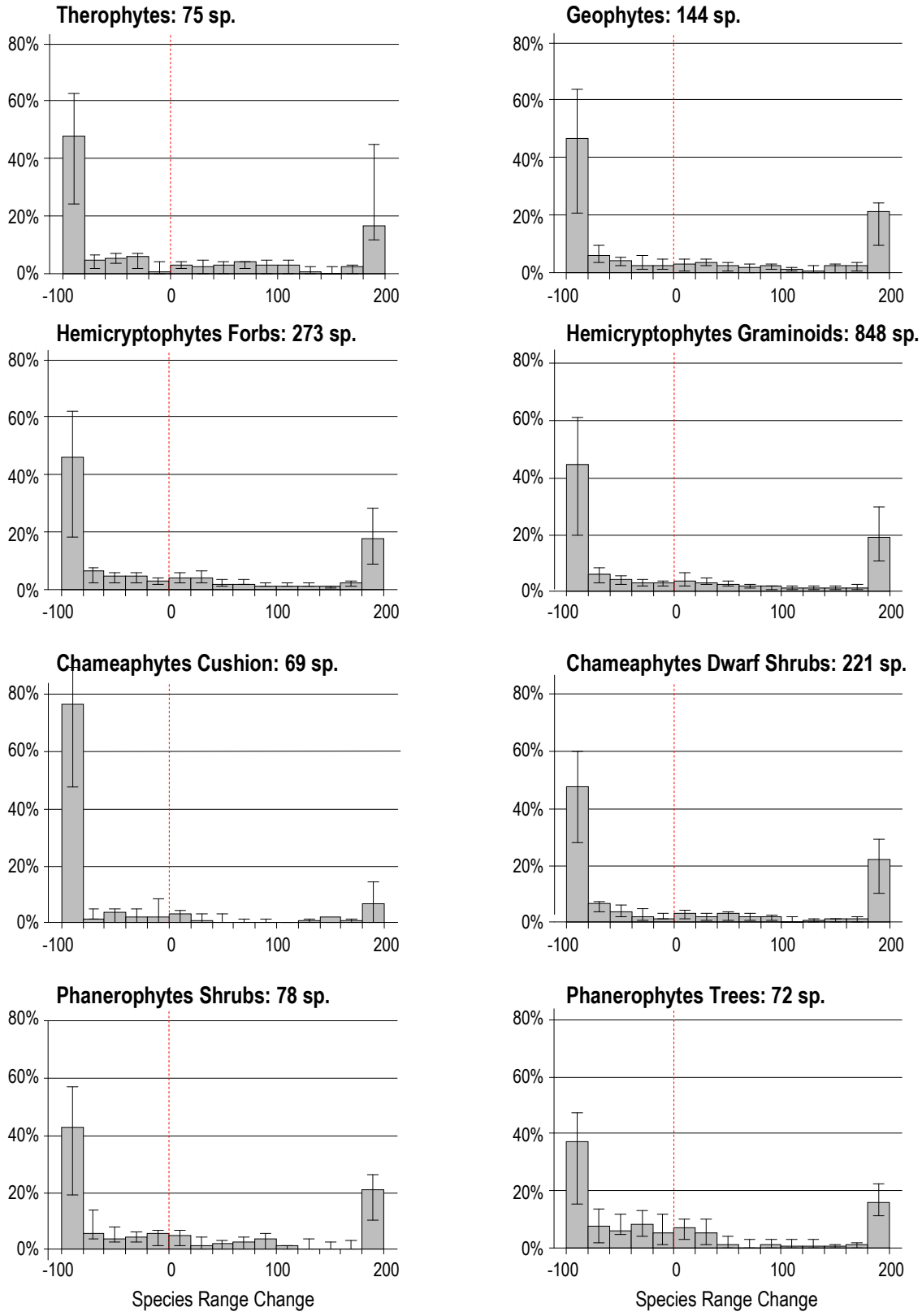


Figure S7.2. Projected SRC for the year 2100 under A1FI climate change scenarios for the different vegetative growth forms of species.

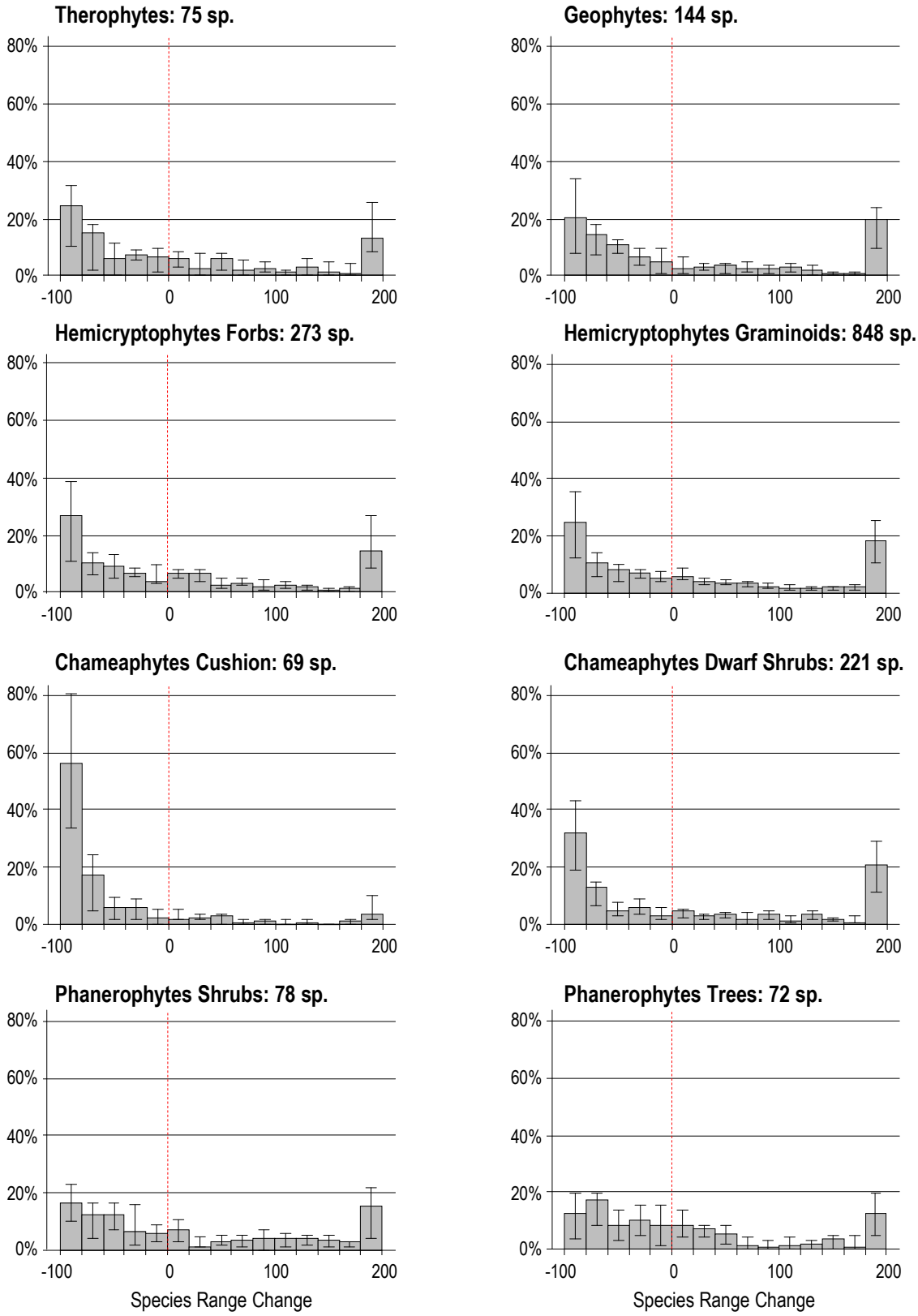


Figure S7.3. Projected SRC for the year 2100 under B1 climate change scenarios for the different vegetative growth forms of species.

Supplementary material S8: Linear and logistic regressions of species extinction rates against study area elevation range and species distribution along the elevation gradient.

Given that the study areas and datasets representing each mountain range have different size, elevation gradient and species pools (Table S8.1), rates of species projected to become locally extinct or importantly reduced cannot be compared directly between them. The two most important bias factors that could be identified to introduce "artificial" variation in species threat levels between study areas are (Fig. S8.1):

1. Study area elevation range. Study areas with greater elevation range are expected to have less species extinctions because they offer more room for species to migrate upwards.
2. Position of species along the altitudinal gradient of the study area, i.e. species located towards the bottom of the study area are expected to be less threatened because they are provided with the opportunity to migrate upwards. This information was summarized in a "species elevation index".

The "species elevation index" for each study area was derived as follows: First, each study area was divided into 10 equal altitudinal slices having each an elevation range of one tenth of the study area's elevation range. These elevation slices received a value from "1", for the top-most slice, to "10", the lowest elevation slice.

Each species was then associated to the altitudinal slice in which the mean of its observed occurrences was falling into (histograms of species occurrence frequency along the elevation gradient were visually checked to ensure that no species had a bi-modal distribution). A species falling into the highest elevation slice was thus awarded a ranking of "1" and a species falling into the lowest elevation slice a ranking of 10.

Finally, to derive the species elevation index, the rankings for all species in a given study area were averaged (Table S8.1). The smaller the index (i.e., the closer to 1), the closer the species' spatial distribution is to the top of the study area, the larger (i.e., the closer to 10), the more the species occupy the lower parts of the study area's elevation gradient.

Table S8.1: Species and physical attributes of the different dataset. * = these study areas were not used in the regression analyses due to their small number of species.

Study Area name	Elevation Range	Species Position along elevation gradient
Austrian Alps	1767	3.72
Carpathians	2250	2.15
French Alps 1	4785	8.56
French Alps 2	1695	4.61
Italian Apennines *	1282	2.90
Norwegian Scandes	2443	5.98
Pyrenees 1	2734	6.80
Pyrenees 2 *	2395	4.40
Scottish Highlands	1095	4.89
Swiss inner Alps 1	3085	7.25
Swiss inner Alps 2	1610	7.39
Swiss Western Alps	2749	5.97

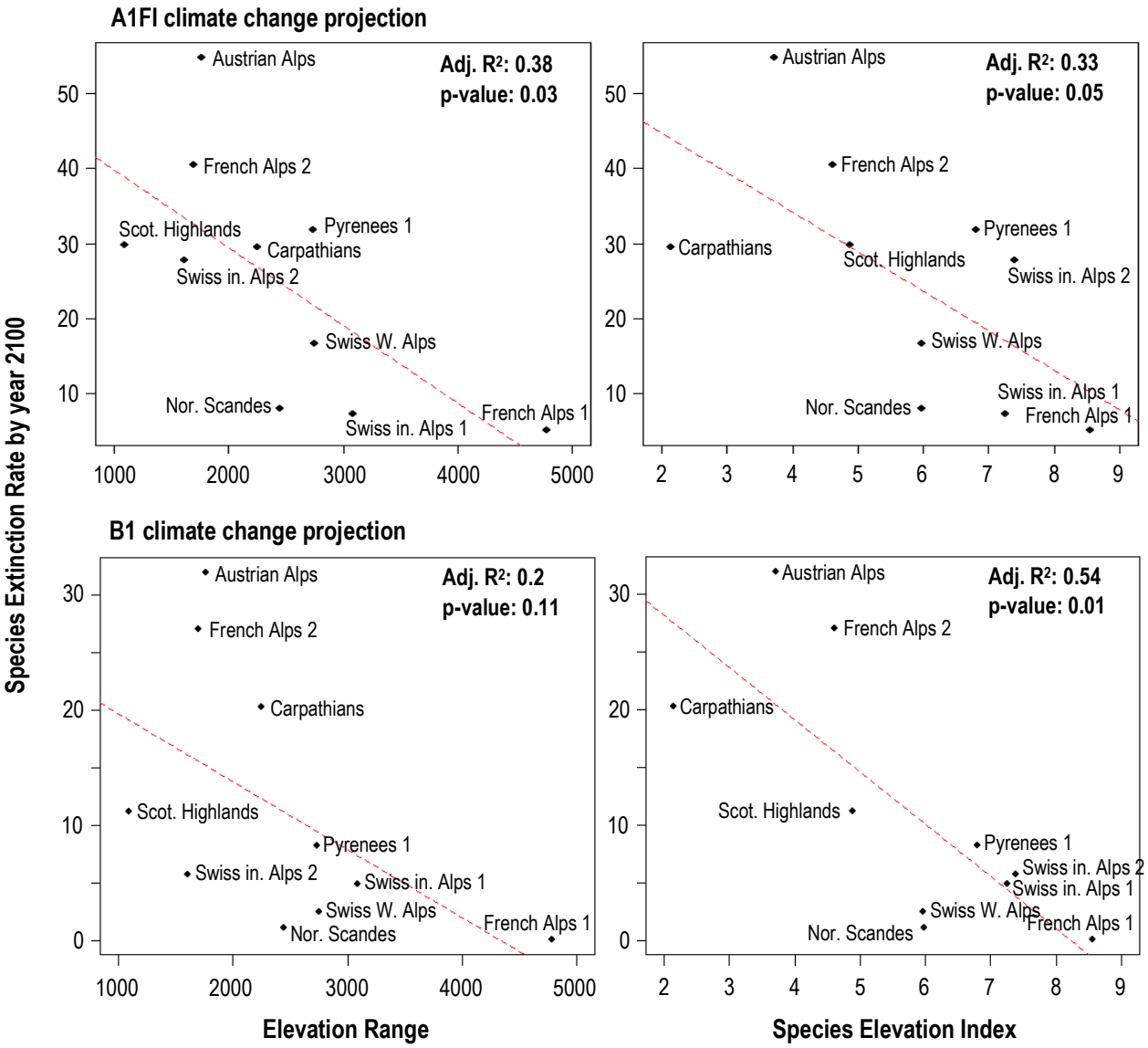


Figure S8.1. Species extinction rate as a function of a study area's elevation range (left panels) or species elevation index (position of species along the altitudinal gradient) under A1FI (upper panels) and B1 (lower panels) climate change projections. The dotted lines represents the regression line of a linear model. Adjusted deviance and p-value associated with the regression are indicated in the upper right corner of each plot.

Logistic Regressions

We fitted binomial logistic regressions (generalized linear models) with logit link ($Y = e^{LP}/(e^{LP} + 1)$, where LP is the linear predictor) to respectively relate the rate of species projected to become extinct (100% decrease in potential distribution) or importantly reduced (>80 decrease in potential distribution) to a study area's elevation range and average species elevation index (Table S8.2). The Italian Apennines and the Pyrenees 2 datasets were left out of this analysis as they had too little species.

Table S8.2. Coefficients and null, explained and residual deviance of the calibrated GLMs. Estimates are given in the space of the linear predictor.

A1Fl Extinction rate (Y) ~ Elevation range (X1) + Species Elevation Index (X2)

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.078e+00  1.614e-01   6.679 2.41e-11 ***
X1          -5.620e-04  6.461e-05  -8.699  < 2e-16 ***
X2          -1.659e-01  2.723e-02  -6.092 1.11e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 389.09  on 9  degrees of freedom
Residual deviance: 161.16  on 7  degrees of freedom
AIC: 223.21
```

B1 Extinction rate (Y) ~ Elevation range (X1) + Species Elevation Index (X2)

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.2226161  0.2489385   4.911 9.05e-07 ***
X1          -0.0006189  0.0001021  -6.063 1.33e-09 ***
X2          -0.3853059  0.0364812 -10.562  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 350.24  on 9  degrees of freedom
Residual deviance: 112.93  on 7  degrees of freedom
AIC: 162.84
```

A1Fl Rate of importantly reduced species (Y) ~ Elevation range (X1) + Species Elevation Index (X2)

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.384e-01  1.261e-01   5.061 4.16e-07 ***
X1          -2.836e-04  4.759e-05  -5.958 2.55e-09 ***
X2          -7.834e-03  2.498e-02  -0.314    0.754
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 237.25  on 9  degrees of freedom
Residual deviance: 178.43  on 7  degrees of freedom
AIC: 245.13
```

B1 Rate of importantly reduced species (Y) ~ Elevation range (X1) + Species Elevation Index (X2)

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  9.612e-01  1.452e-01   6.622 3.54e-11 ***
X1          -3.440e-04  5.766e-05  -5.966 2.43e-09 ***
X2          -1.862e-01  2.630e-02  -7.080 1.44e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 391.93  on 9  degrees of freedom
Residual deviance: 207.74  on 7  degrees of freedom
AIC: 271.34
```

Synthesis and Perspectives

S*pecies distribution models for conservation-related applications*

The first major contribution of this thesis has been to contribute to evidence the potential of species distribution models (SDMs) for conservation-related applications. In chapter 1.1 – Engler et al. 2004, the model developed for the rare *Eryngium alpinum* (Alpine Eryngo or Queen of the Alps) gives an indication of how the projections of a habitat suitability model can help discovering new populations or rare species, although without providing a rigorous demonstration. Chapter 1.2 (along with Guisan et al. 2006 – Annex 1) and Chapter 1.3 further develop the methodology of model-based sampling and comprehensively demonstrate its efficiency through both simulations and field work.

In parallel to the work that is presented in this thesis, other research groups have also developed and tested the model-based sampling approach. Edwards et al. (2005) applied it to detect rare lichen species, and obtained a two- to five-fold increase in detection rate as compared to random sampling. More recently, Aitken et al. (2007) also concluded that model-based sampling enhances the efficiency of new population discovery and Williams et al. (2009) found new populations for four out of six rare species when using model-based sampling in a study area of northern California. Taken together with the results presented in this thesis, there is thus strong evidence that model-based sampling has the potential to become a useful tool for practical applications in conservation biology.

Obviously, the usefulness of the model-based sampling approach depends on the spatial prevalence of the target species: the rarer the species (rare being defined here as having low spatial prevalence), the more powerful the approach will be in comparison of other sampling strategies. If the target species is very common, then model-based sampling will become an unnecessary complication and a more traditional sampling strategy, such as a random-stratified sampling based on a combination of major environmental gradients, might be preferable.

How are pseudo-absences best selected?

Another aspect presented in this thesis is the selection of pseudo-absences to overcome limitations associated with presence-only data (Chapter 1.1 – Engler et al. 2004).

As noted by Rushton et al. (2004) in their editorial accompanying the publication of the manuscript, creating a first model to derive pseudo-absences data that are to be used in a second model introduces some bias into the data and thus requires careful consideration. What this means is that, if the first model, based on an ENFA – or any other presence-only modeling technique – is of bad quality and wrongly discriminates between suitable and unsuitable habitats, then the generated pseudo-absences are likely to further enforce this bias and the final model will be worse than if it was based on random pseudo-absences.

Another criticism of this approach is that it does not address spatial bias in occurrence data, which is frequent in records from herbaria and natural history museums (Phillips et al. 2009). If occurrence data are spatially biased, the resulting map of potentially suitable habitat is likely to picture sampling effort rather than genuine habitat suitability, and selecting the pseudo-absences in areas with low suitability will only make matter worse by further increasing the initial bias. As a result, the approach that is presented in chapter 1.1 is certainly double-edged: while it can improve projections if the initial presence-only data is of quality, it can also introduce further bias into already biased data.

Because most available datasets are lacking reliable absence data, the question of how to best generate pseudo-absences has received considerable attention over the recent years (Engler et al. 2004 has been cited more than 140 times). Most studies have adopted the strategy of weighting the selection of pseudo-absences towards *a-priori* low-suitability habitats (e.g. Zaniewski et al. 2002, Lütolf et al. 2006), as we have done in chapter 1.1. But alternatives methods for the weighting of pseudo-absences have also been developed. An interesting example is provided by the work of Zarnetske et al. (2007), who generate their pseudo-absences only within the area that is projected as suitable for the species. The rationale behind this approach is that, by selecting pseudo-absences from those areas that are favorable to the species, the model will be able to more finely distinguish between suitable and highly suitable habitats.

Some studies have also simply generated a very large number of pseudo-absences completely at random and gave them a lower weight during model calibration (e.g. Guisan et al. 2006). In this case, the idea is that even if some of the pseudo-absence points are located in environmental conditions that are actually suitable for the target species, their number and weight will be much lower than that of presence records and they will thus have a negligible effect during model calibration. While this method does not correct for initial sampling bias in presence records, it at least avoids the risk of including additional bias when selecting the pseudo-absences.

In a recent contribution by Chefaoui and Lobo (2008), these different approaches to generate pseudo-absence are compared. The results reveal that there is not one best pseudo-absence selection strategy but rather that different selection strategies result in different types of projections: the further the pseudo-absences are selected from the true suitable habitat, the closer the projection will be to the fundamental niche of the species. Conversely, if the pseudo-absences are weighted in favor of suitable locations, then the projections will be closer to the realized niche of the species. Thus, according to the results of Chefaoui and Lobo (2008), the optimal pseudo-absence generation strategy largely depends on what the modeling is supposed to achieve, i.e. what is the intended use of the generated habitat suitability map.

Finally, a promising approach is proposed in a recent paper by Phillips et al. (2009), who suggest weighting the selection of pseudo-absences towards those areas where the sampling effort has been allocated to the collection of the occurrence records. This approach is particularly interesting as it allows for the removal of potential spatial bias in the presence records.

Assessing the potential impacts of climate change on vegetation

The second major achievement of this thesis has been to contribute to improve projections of potential climate change impact of species distributions. This was accomplished by (i) the development of a model that enables the implementation of dispersal limitations and basic population dynamics into classical species distribution models (Chapter 2.1 – Engler and Guisan 2009), and of a simplified classification of species into "dispersal types" that allows to derive dispersal distance estimates for a large number of species (Chapter 2.2 – Vittoz and Engler 2007). (ii) The assessment of the importance of accounting for dispersal limitations in projections made for mountain areas (Chapter 2.3 – Engler et al. 2009). (iii) The assessment of

potential climate change impacts at fine scale over a large number of mountain regions distributed over western and central Europe (Chapter 3.1).

Local fine scale vs. global coarse scale models

In chapter 3.1 a fine scale assessment of potential climate change impacts on mountain vegetation across major European mountain ranges is presented. This study highlights that some mountain ranges (the Pyrenees and the Austrian Alps) are expected to be more affected by climate change than others (Norway and the Scottish Highlands), thereby corroborating findings of Thuiller et al. (2005) who obtained a similar pattern when working at much coarser scale. The comparison of the potential impact of climate change between the different study areas also emphasizes the importance of study area elevation range and species pool in projected threat levels. In example, a smaller elevation range will result in higher projected extinction rates. Therefore, if the limits of a study area are artificial (i.e. the study area does not encompass the entire available altitudinal gradient), then the projected threat levels need to be given careful consideration. Finally, the comparison of the results we obtained in this study with those of Thuiller et al. (2005) indicate that the threat levels projected through fine scale modeling are less severe than those derived from coarse scale models. In other words, this result suggests that some species could persist in small refugias that are not detected by coarse scale models. While this latter result corroborates with earlier findings (Randin et al. 2009 – Annex 3), others have come to opposite conclusions (i.e. fine scale models predict less species persistence; Trivedi et al. 2008). The issue of divergence between global coarse scale vs. local fine scale and how to reconcile them thus remains open and will need further examination.

How important is accounting for dispersal limitations?

The importance of dispersal limitations in projections of future species distribution under climate change has been given special attention in this thesis (Chapter 2). The answer to this question is not clear cut: it all depends on what is assessed, at which scale and where it is assessed.

When assessing species threat levels (i.e. rates of species projected to become extinct of highly reduced in distribution) in our study area that is mountainous and of limited size (local or regional assessment), considering dispersal as unlimited provided satisfying results that were close to those obtained from the simulations that accounted for dispersal limitations . Although this result should be put at test in other study areas and with different species pools than ours, it is probably reasonable to think that it can be extrapolated to similarly sized study areas that are also located in mountain areas.

However, as I stated above, "it all depends", and therefore a number of remarks need to be made to accompany the above paragraph: Most importantly, the relative unimportance of taking into account dispersal limitations only applies to the assessment of species threat levels. If other elements are to be assessed, then dispersal limitations need to be accounted for (Engler and Guisan 2009 – Chapter 2.1; Midgley et al. 2006). Furthermore, this result is only valid for projections that are made on a local to regional scale. Thus the relative unimportance of dispersal limitations does not apply to threat levels in general but only to *local* threat levels (i.e. a species can become extinct from a certain region without disappearing from the entire planet). Finally, this result is only valid for mountain regions and results in flatter areas are expected to be very different (i.e. dispersal limited simulations could be closer to no-dispersal than to unlimited dispersal, see e.g. Midgley et al. 2006).

While the above remarks must be kept in mind, the finding that, when estimating local threat levels (e.g. projected local extinction rate) of vegetation from climate change, the results obtained from projections assuming unlimited dispersal offer a reasonable approximation, remains of importance. Especially in the light of the important additional effort that is required to account for dispersal limitations. This result also *a-posteriori* strengthens the projections from previous studies that have investigated the potential impacts of climate change in mountains areas (e.g. Guisan and Theurillat 2000, Dirnböck et al. 2003, Randin et al. 2009) but have considered dispersal as unlimited. However, if working at larger geographical scales or on one species in particular, then running a sensitivity analysis to quantify the effects of dispersal limitations is to be strongly encouraged.

One of the strengths of the MIGCLIM model developed in chapter 2.1 (see also Annex 5) is that is can be associated with any modeling technique and is flexible in its parameters. Another advantage of MIGCLIM is that is can give the potential distribution of a species in a continuous fashion over time (i.e., it does not only give the potential distribution of the target species at the end of the simulation but also at regular, user defined, intervals during the simulation). This is of importance, as one is not only interested to know how many species might become threatened due to climate change, but also when these species are projected to become importantly threatened (or more generally how their potential distribution is evolving over time).

From a more pragmatic perspective, the practical success of MIGCLIM (or any other model for that matter) will depend on how user-friendly and easy to use it is. Ideally, a model like MIGCLIM should therefore be coupled tightly with an existing modeling package (e.g. BIOMOD, Thuiller et al. 2009 – Annex 2). Another practical issue that might refrain potential users is the calibration of the model's parameters. While MIGCLIM remains a relatively simple model in terms of number of parameters as compared other existing models (e.g. Dullinger et al. 2004, Scheller et al. 2007), calibrating even its limited number of parameters can prove difficult, especially if it needs to be done for several species. Furthermore, the uncertainty in parameters needs to be assessed (Higgins et al. 2003) to avoid that a more complex model simply provides a false impression of accuracy rather than a true improvement in predictive ability.

Nevertheless, even if very accurate values for a given parameter cannot be achieved, models such as MIGCLIM can always be used to run sensitivity analyses on different parameters. One parameter that is particularly illustrative in this regard are long distance dispersal (LDD) events, that allow the movement of seeds over unusually large distances and often using unusual – and thus difficult to model – means of dispersal (see Nathan et al. 2008 for a review). Despite the important body of literature dedicated to seed dispersal, the importance of long-distance dispersal in plant migration remains a hotly debated topic (Pearson, 2006). On one hand, several studies have shown its importance in determining plant dispersal rates (Higgins and Richardson 1999, Pearson and Dawson 2005, Soons and Ozinga 2005), and especially for explaining fast plant migrations in the past (Cain et al. 1998, Clark 1998, Nathan and Muller-Landau 2000). On the other hand, late-glacial refugia and molecular data indicate that migration rates might not have been as high as initially thought (see Pearson 2006).

Although methods to estimate LDD events exist (Nathan et al. 2003) and significant improvements have been made in LDD modeling (e.g. Tackenberg 2003, Soons et al. 2004, Nathan et al. 2005), accurately integrating them into spatially explicit models of species distribution and dispersal remains difficult. In fact, some have argued that accurate

predictions of LDD events might not even be possible due to their inherent unpredictability (Clark et al. 2003).

In the absence a of clear consensus on the importance of these long distance dispersal events, the use of models takes all its importance as it allows to test the effect of different hypothesis (in this particular case testing whether the inclusions of these long distance dispersal events has a significant effect on projections). For instance, repeating a same simulation (with a stochastic LDD event component) would allow to quantify the uncertainty related to LDD in both space and time. Adding such assessment of uncertainty is certainly a direction that should be taken and will increase the confidence that can be awarded to our model projections. This is especially important if these models are to be used for practical implementation of climate change mitigation policies.

References

Aitken M., Roberts D.W. and Shultz L.M., 2007. Modeling distributions of rare plants in the Great Basin, western North America. *Western North American Naturalist*, **67**(1), 26-38.

Cain M.L., Damman H. and Muir A., 1998. Seed Dispersal and the Holocene Migration of Woodland Herbs. *Ecological Monographs*, **68**(3), 325-347.

Chefaoui R.M. and Lobo J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, **210**(4), 478-486.

Clark J.S., 1998. Why Trees Migrate So Fast: Confronting Theory with Dispersal Biology and the Paleorecords. *The American Naturalist*, **152**(2), 204-224.

Clark J.S., Lewis M., McLachlan J.S. and HilleRisLambers J., 2003. Estimation population spread: what can we forecast and how well? *Ecology*, **84**(8), 1979-1988.

Dirnböck T., Dullinger S. and Grabherr G., 2003. A regional impact assessment of climate and land-use change on alpine vegetation. *Journal of Biogeography*, **30**(3), 401-417.

Dullinger S., Dirnbock T. and Grabherr G., 2004. Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. *Journal of Ecology*, **92**(2), 241-252.

Edwards T.C., Cutler D.R., Zimmermann N.E., Geiser L. and Alegria J., 2005. Model-based stratifications for enhancing the detection of rare ecoligical events. *Ecology*, **86**(5), 1081-1090.

Engler R., Guisan A. and Rechsteiner L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41**(2), 263-274.

Engler R. and Guisan A., 2009. MIGCLIM: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, **15**(4), 590–601.

Engler R., Randin C.F., Vittoz P., Czáka T., Beniston M., Zimmermann N.E. and Guisan A., 2009. Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? *Ecography*, **32**(1), 34-45.

Guisan A. and Theurillat J.-P., 2000. Assessing alpine plant vulnerability to climate change: A modeling perspective. *Integrated Assessment*, **1**, 307-320.

Guisan A., Broennimann O., Engler R., Vust M., Yoccoz N.G., Lehmann A. and Zimmermann N.E., 2006. Using niche-based models to improve the sampling of rare species. *Conservation Biology*, **20**(2), 501-511.

Higgins S.I. and Richardson D.M., 1999. Predicting Plant Migration Rates in a Changing World: The Role of Long-Distance Dispersal. *American Naturalist*, **153**, 464-475.

Higgins S.I., Clark J.S., Nathan R., Hovestadt T., Schurr F., Fragoso J.M.V., Aguiar M.R., Ribbens E. and Lavorel S., 2003. Forecasting plant migration rates: managing uncertainty for risk assessment. *Journal of Ecology*, **91**(3), 341-347.

Lütolf M., Kienast F. and Guisan A., 2006. The ghost of past species occurrence: improving species distribution models for presence-only data. *Journal of Applied Ecology*, **43**(4), 802-815.

Midgley G.F., Hughes G.O., Thuiller W. and Rebelo A.G., 2006. Migration rate limitations on climate change-induced range shifts in Cape Proteaceae. *Diversity and Distributions*, **12**, 555-562.

Nathan R. and Muller-Landau H.C., 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology and Evolution*, **15**(7), 278-285.

Nathan R., Perry G., Cronin J.T., Strand A.E. and Cain M.L., 2003. Methods for estimating long-distance dispersal. *Oikos*, **103**, 261–273.

Nathan R., 2005. Long-distance dispersal research: building a network of yellow brick roads. *Diversity and Distributions*, **11**, 125–130.

Nathan R., Schurr F.M., Spiegel O., Steinitz O., Trakhtenbrot A. and Tsoar A., 2008. Mechanisms of long-distance seed dispersal. *Trends in Ecology and Evolution*, **23**(11), 638-647.

Pearson R.G. and Dawson T.P., 2005. Long-distance plant dispersal and habitat fragmentation: identifying conservation targetsfor spatial landscape planning under climate change. *Biological Conservation*, **123**, 389-401.

Pearson R.G., 2006. Climate change and the migration capacity of species. *Trends in Ecology and Evolution*, **21**(3), 111-113.

Phillips S.J., Dudík M., Elith J., Graham C.H., Lehmann A., Leathwick J. and Ferrier S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**(1), 181-197.

Randin C.F., Engler R., Normand S., Zappa M., Zimmermann N.E., Pearman P.B., Vittoz P., Thuiller W. and Guisan A., 2009. Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology*, **15**, 1557-1569.

Rushton S.P., Ormerod S.J. and Kerby G., 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology*, **41**(2), 193-200.

Scheller R.M., Domingo J.B., Sturtevant B.R., Williams J.S., Rudy A., Gustafson E.J. and Mladenoff D.J., 2007. Design, development, and application of LANDIS-II, a spatial landscape simulation model with flexible temporal and spatial resolution. *Ecological Modelling*, **201**(3-4), 409-419.

Soons M.B., Heil G.W., Nathan R. and Katul G.G., 2004. Determinants of long-distance seed dispersal by wind in grasslands. *Ecology*, **85**(11), 3056–3068.

Soons M.B. and Ozinga W.A., 2005. How important is long-distance seed dispersal for the regional survival of plant species? *Diversity and Distributions*, **11**, 165–172.

Tackenberg O., 2003. Modeling long-distance dispersal of plant diaspores by wind. *Ecological Monographs*, **73**(2), 173–189.

Thuiller W., Lavorel S., Araujo M.B., Sykes M.T. and Prentice I.C., 2005. Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(23), 8245-8250.

Thuiller W., Lafourcade B., Engler R. and Araújo M.B., 2009. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369-373.

Trivedi M.R., Berry P.M., Morecroft M.D. and Dawson T.P., 2008. Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. *Global Change Biology*, **14**(5), 1089-1103.

Vittoz P. and Engler R., 2007. Seed dispersal distances: a typology based on dispersal modes and plant traits. *Botanica Helvetica*, **117**(2), 109-124.

Williams J.N., Changwan S., Thorne J., Nelson J.K., Erwin S., O'Brien J.M. and Schwartz M.W., 2009. Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, **15**, 565–576.

Zaniewski A.E., Lehmann A. and Overton J.M.C., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, **157**(2-3), 261-280.

Zarnetske P.L., Edwards T.C. and Moisen G.G., 2007. Habitat classification modeling with incomplete data: Pushing the habitat envelope. *Ecological Applications*, **17**(6), 1714-1726.

Annexes

A1 Using niche-based models to improve the sampling of rare species.

A2 BIOMOD - a platform for ensemble forecasting of species distributions.

A3 Climate change and plant distribution: local models predict high-elevation persistence.

A4 Prediction of plant species distributions across six millennia.

A5 MIGCLIM User Guide.

A 1

Using niche-based models to improve the sampling of rare species.

Guisan A., Broennimann O., Engler R., Vust M., Yoccoz N.G., Lehmann A. and Zimmermann N.E, 2006. *Conservation Biology*, **20**, 501-511.

Using Niche-Based Models to Improve the Sampling of Rare Species

ANTOINE GUISAN,*†† OLIVIER BROENNIMANN,* ROBIN ENGLER,* MATHIAS VUST,* NIGEL G. YOCCOZ,† ANTHONY LEHMANN,§ AND NIKLAUS E. ZIMMERMANN‡

*University of Lausanne, Department of Ecology and Evolution (DEE), Laboratory for Conservation Biology (LBC), Biology Building, CH-1015 Lausanne, Switzerland
†Institute of Biology, University of Tromsø, 9037 Tromsø, Norway
‡Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
§Swiss Center for Faunal Cartography, Terreaux 14, CH-2000 Neuchâtel, Switzerland

Abstract: *Because data on rare species usually are sparse, it is important to have efficient ways to sample additional data. Traditional sampling approaches are of limited value for rare species because a very large proportion of randomly chosen sampling sites are unlikely to shelter the species. For these species, spatial predictions from niche-based distribution models can be used to stratify the sampling and increase sampling efficiency. New data sampled are then used to improve the initial model. Applying this approach repeatedly is an adaptive process that may allow increasing the number of new occurrences found. We illustrate the approach with a case study of a rare and endangered plant species in Switzerland and a simulation experiment. Our field survey confirmed that the method helps in the discovery of new populations of the target species in remote areas where the predicted habitat suitability is high. In our simulations the model-based approach provided a significant improvement (by a factor of 1.8 to 4 times, depending on the measure) over simple random sampling. In terms of cost this approach may save up to 70% of the time spent in the field.*

Key Words: efficiency, endangered species, *Eryngium alpinum*, habitat suitability maps, population discovery, predicted species distribution, prospective sampling

Utilización de Modelos Basados en Nichos para Mejorar el Muestreo de Especies Raras

Resumen: *Debido a que los datos sobre especies raras generalmente son escasos, es importante contar con formas eficientes para obtener datos adicionales. Los métodos de muestreo tradicionales tienen valor limitado para especies raras porque en una gran proporción de sitios de muestreo seleccionados al azar es poco probable que se encuentre a la especie. Para estas especies, las predicciones espaciales de modelos de distribución basados en nicho pueden ser utilizadas para estratificar el muestreo e incrementar la eficiencia de muestreo. Los nuevos datos obtenidos luego son utilizados para mejorar el modelo inicial. La aplicación repetida de este método es un proceso adaptativo que puede permitir el incremento del número de ocurrencias nuevas. Ilustramos el método con un estudio de caso de una especie de planta rara y en peligro en Suiza y un experimento de simulación., Nuestro trabajo de campo confirmó que el método ayuda al descubrimiento de nuevas poblaciones de la especie en áreas remotas en las que la adecuación pronosticada del hábitat es alta. En nuestras simulaciones, el método basado en modelos aportó un mejoramiento significativo (por un factor de 1.8 a 4 veces, dependiendo de la medida) del muestreo aleatorio simple. En términos de costos, este método puede ahorrar hasta 70% del tiempo de trabajo de campo.*

Palabras Clave: descubrimiento de población, eficiencia, distribución pronosticada de especies, *Eryngium alpinum*, especies en peligro, mapas de adecuación del hábitat, muestreo prospectivo

††*email antoine.guisan@unil.ch*
Paper submitted February 23, 2005; revised manuscript accepted June 8, 2005.

Conservation Biology Volume 20, No. 2, 501–511

DOI: 10.1111/j.1523-1739.2006.00354.x

Introduction

Monitoring populations of rare and endangered species has become a priority for most conservation agencies. It provides the major source of data for updating World Conservation Union and national red lists (Lamoreux et al. 2003), the main use of which is setting the long-term goals of conservation programs, such as helping identify biotopes or areas particularly in need of protection (Prendergast et al. 1993). Although precise quantitative and qualitative criteria are used to assign a particular status to a species (IUCN/SSC 2001), such decisions depend on the way the target species populations are monitored and on the sampling design used for data collection. A range of sampling schemes can be used to estimate the state of the system and its rate of change, the most efficient of which usually combine repeated (random) samples among existing sites with additional independent random samples at unvisited locations (Yoccoz et al. 2001; Stauffer et al. 2002).

Yet this random component is seldom considered in conservation surveys of rare species, although it is fundamental for ensuring unbiased population estimates and calculating unbiased estimates of distributional and population states. For example, remote populations may differ from accessible ones with respect to important environmental and management variables. In the context of distribution modeling, bias is thus likely to affect the quantification of the realized niche of a species and therefore the assessments that could be based on it. These include assessing the impact of climate change on species distribution and associated risks of extinction or searching for reintroduction sites for a species.

When species are rare, standard sampling methods such as simple or stratified random sampling, based on a simple combination of the main environmental gradients, can be highly inefficient (Rushton et al. 2004). For instance, in a sample of 550 plots surveyed in a random-stratified way based on the elevation, slope, and aspect of the plot during two consecutive summers in the Swiss Alps (704.2 km^2), not one occurrence of the rare and endangered plant species *Eryngium alpinum* L. was recorded. This was despite the species being easily detectable if present and independent records of the species existing in the area within similar vegetation types. Directing the sampling to ensure inclusion of units with a higher probability of hosting the species is thus a desirable approach for increasing survey efficiency and reducing sampling costs.

Predictions from niche-based models of species distribution (Guisan & Zimmermann 2000) are promising tools in this respect (Côté & Reynolds 2002; Edwards et al. 2005) as a way to improve the sampling of species of conservation interest. Although population viability analyses have long been used in rare species management (Brook et al. 2000), spatially explicit, predictive, habitat distribution models have only recently been used in conservation biology (Vaughan & Ormerod 2003; Rushton et al. 2004). The majority of predictive models published in the literature were developed for common plant and animal species or for biodiversity. To date, relatively few successful applications of this approach have been published for rare and endangered plant species (but see Miller 1986; Elith & Burgman 2002; Engler et al. 2004), although reliable spatial predictions are essential for species of great conservation interest. Paucity of data, spatial inaccuracy, and lack of valid absences are the main reasons identified for this shortcoming (Engler et al. 2004).

Recent methodological progress (Guisan & Thuiller 2005), such as improved predictive algorithms and more causal environmental predictors at a better spatial resolution, has made predictive models more reliable and applicable to the sampling of rare and endangered species (Ferrier 2002). Here, we propose a procedure for using niche-based models of species distribution to direct a random-stratified prospect of the study area and improve the sampling of rare and endangered species. We illustrate the approach with the rare species *E. alpinum* by using data taken from a modeling study of endangered plant species in Switzerland (O.B., unpublished). We also tested the approach through simulations. Finally, we discuss the conditions for the successful application of this approach, highlight some of its main limitations, and make suggestions for future research.

Methods

A General Procedure for Model-Based Sampling of Species Distribution

The general procedure for model-based sampling of species distribution is pictured in Fig. 1. A first model is

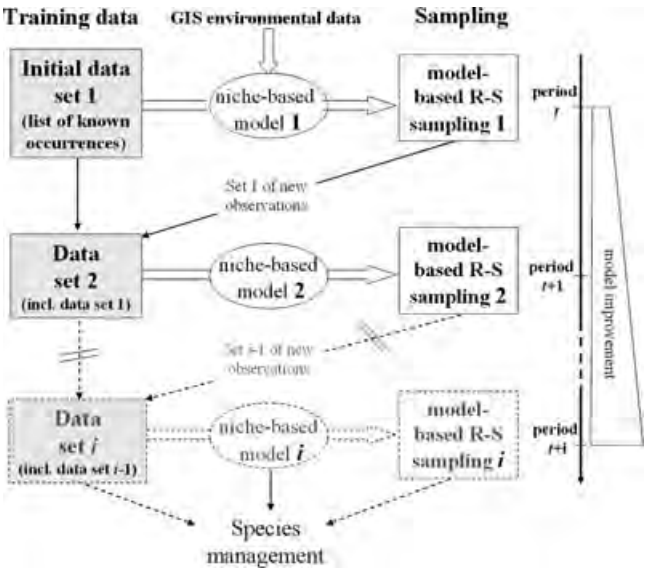


Figure 1. Analytical procedure illustrating the iterative model-based sampling process.

fitted with available data. These data include information on monitored populations or records of species that are typically extracted from a computer data bank, such as those developed in botanical conservatories, museums, or biological data centers (Graham et al. 2004). Spatial predictions derived from this initial model are then used to stratify the field sampling. Data gathered during this first sampling campaign serve to update the training data set and fit improved models that are used to direct the next sampling stage, and so on iteratively over several field seasons (Fig. 1). The higher prediction strata are expected to reveal those occurrences of suitable environmental conditions for the species where undetected or newly established populations, not recorded in the initial data set, might prevail. This iterative process can be repeated during the same field season or over consecutive seasons, and its success is mainly measured by the number of new occurrences found at each stage.

Study Area and Species Data

The study area is the same as in Engler et al. (2004). We illustrate the proposed approach with *E. alpinum*, commonly called alpine eryngo, an emblematic species endangered in most parts of the calcareous European Alps. It grows mainly on deep and moist soils with intermediate to high level of nutrients, preferably on steep slopes where screes of large blocks occur and solar radiation is rather high. The exact reasons for its recent decline are still unknown, but possible threats could be picking and changes in pastoral practices (Engler et al. 2004).

Original species data on *E. alpinum* were provided by the Swiss Floristic Network (CRSF) in Geneva (data can be ordered from http://www.crsf.ch). These records are usually provided by various volunteers and professionals, and thus they share the same idiosyncrasies as those in other natural history collections (Graham et al. 2004) except that, in our case, information on the positional accuracy of each record is also provided. Only data with a minimum horizontal accuracy of 25 m were retained for the analyses, to match the resolution of environmental data layers used. Remaining data were used for visual evaluation of the prediction maps. Because only presence observations were initially available in the database, fitting the initial model required the generation of pseudoabsences. This was based on occurrences of 11 other rare species. With easily detectable and well-known rare species, like *E. alpinum*, this method of generating pseudoabsences works well because providers of rare species observations usually have a good knowledge of most other rare species. Thus, for each single observation of a rare species one can confidently assume the absence of all other possible rare species that usually share the same type of habitat.

A total of 92 presences and 2380 absences were used to develop the initial predictive model. Following the recommendations of Manel et al. (2001), we weighted the absences in the generalized linear models (GLMs) to ensure an equal prevalence (0.5) between presences and absences. This was achieved by providing a vector of weights with "1" for presences and "92/2380" for absences; so the total weight of absences was equal to the total weight (sum) of presences.

Environmental Predictors

We selected quantitative predictors to reflect the main biophysical gradients with a recognized, physiological influence on plants. We generated maps of monthly precipitation sums and monthly temperature averages from the Swiss meteorological network (normals 1961–1990) and a 25-m digital elevation model (DEM) and used them to derive predictors with a more causal effect on the fitness and survival of species. Topographic positions and slope angle were additionally derived from the DEM to express microclimatic corrections and disturbance effects. Table 1 lists all the predictors that were used to fit the models. Details on these predictors can be found in Zimmermann and Kienast (1999).

Table 1. Quantitative geographic information system predictor variables used to fit the models for *Eryngium alpinum*.

Variables	*Description*
Seasonal indices of precipitation	obtained by a principal component analysis on monthly precipitation maps; axis 1 is the yearly sum of precipitation, whereas axes 2 and 3 express two seasonal variations trends
Monthly average of potential daily global radiation in March and July	calculated from the digital elevation model (DEM) as the sum of direct and diffuse radiation
Number of days of precipitation per growing season	calculated from a regionalized regression model of the annual number of days with rainfall of more than 1 mm on elevation
Degree-days of growing season	calculated as the sum of interpolated temperatures above the threshold of 3° C
Number of frost days	defined as a sudden drop of the daily minimum temperature below 2° C preceded by a period of at least 1 day above 3° C; expresses the core 90% of the vegetation period
Topographic position	relative topographic exposure of a pixel compared with its local neighborhood; indicates ridge tops, regular slopes, or valley bottoms
Site water balance	estimate of the water available to plants during a water year, based on precipitation, evapotranspiration, and the soil bucket size derived from the soil suitability map of Switzerland and topographic position
Slope angle	derived from the DEM

Qualitative predictors were broken down into disjunctive classes, which were then used to filter the predictions made by the models. For qualitative filters we used simplified maps of geology and land cover types.

Niche-Based Statistical Models

We used generalized additive models (GAMs; Hastie & Tibshirani 1986) with a binomial distribution and logistic link function to fit the models based on topographic and climatic predictors (hereafter topoclimatic models). Model selection followed a stepwise algorithm based on Akaike's information criterion (AIC) that operates both backward and forward from an initial full model. We used splines as smoothers, with up to four degrees of freedom allowed in the stepwise procedure, as set by default in the program GRASP (Generalized Regression Analysis and Spatial Prediction; Lehmann et al. 2002*b*).

Each topoclimatic model was filtered by the binary variables derived from the qualitative predictors. To be a filter, the species must never occur on any pixel of the corresponding disjunctive class. Typical filters for *E. alpinum* include forested areas and all classes of siliceous bedrock. Although this might have the potential to bias the future sampling, for instance if the original data are biased against some rock or vegetation classes, such a risk is mostly reduced by also sampling outside the suitable areas (see below). Care should thus be taken when using such filters in other situations.

The agreement between predictions and observations was assessed using the standard area-under-the-curve (AUC) measure of a receiver-operating characteristic (ROC) plot (Fielding & Bell 1997) and its cross-validated version (AUC_{CV}). Values of AUC vary between 0.5 for an uninformative, random model and 1 for a model with perfect discrimination. Cross-validation was performed by splitting the data set into five partitions (5-fold CV) and reestimating the model coefficients at each loop.

The minimal predicted area (MPA; Engler et al. 2004) was calculated to complement the other evaluation measures and compare the predicted maps obtained after each loop of the reiterative modeling/field testing process (Fig. 1). It also constituted the primary stratification factor for the field sampling. The MPA is the surface obtained by considering all pixels with predicted values above the probability threshold that still encompasses 100% of the species occurrences (rule of parsimony). The resulting MPA map is binary, with 0 for cells where the habitat is predicted unsuitable and 1 for cells where it is predicted suitable.

To make spatial predictions of species distributions on a large data set possible (64 million pixels at 25-m resolution for Switzerland), we used lookup tables generated from the GAM models. A custom script was used in the ArcView geographic information software (ESRI, Redland, California) to implement the models and calculate the final potential maps.

The entire procedure, except the use of filters, was automated with the GRASP library of S-PLUS (Insightful Corp., Seattle, Washington) functions (Lehmann et al. 2002*b*; GRASP and the ArcView script are available online from http://www.cscf.ch/grasp).

Implementing the Model-Based, Random-Stratified Sampling

The model-based sampling was designed as follows. First, we fitted a model for the species based on topographic and climatic predictors, filtered by binary classes of qualitative predictors and used to predict the species distribution over the study area. Second, we calculated the MPA. Third, we superimposed the MPA map on a relief map and distinguished subareas of similar geological substratum belonging to the same macrotopographic unit (e.g., large massifs) (Fig. 2a). Because of limited sampling resources, only subareas including at least one patch of cells with high habitat suitability for the species (i.e., above the MPA threshold) were considered further for the sampling (Fig. 2a). We visited a random selection of these subareas in the field, according to a standardized field methodology. Hence we did not formally use a probability sample of the whole area, and the results should not be interpreted outside the range of the candidate subareas that define our target population.

We conducted survey walks in visited subareas, following a random trajectory based on the MPA map and designed to cross approximately an equal number of suitable and unsuitable pixels. In an attempt to reduce costs and increase survey efficiency, we opted for survey walks across the probability gradient rather than a strict random-stratified procedure, based on sampling individual pixels randomly in both strata (within and outside MPA), because surveying all of Switzerland was not possible. The full trajectory was recorded using a GPS system (Garmin eTrex Summit, Olathe, Kansas, locational accuracy <10 m). New occurrences, when found, were recorded and georeferenced using the GPS. A series of points without species observations were also sampled at regular intervals (at least 500 m, to avoid spatial autocorrelation) along the recorded trajectory to generate likely absences for the species.

All new and updated presences and absences were used to fit a second, improved model to be used as a next step to redirect the sampling effort (Fig. 1). New observed absences force the model to be more discriminating at locations where the prior model was predicting high suitability values. On the other hand new observations of presences help remove bias from predictions and narrow the species realized niche. A bias may for instance be introduced in the original model if the species observations offered only a partial view of the species niche, which

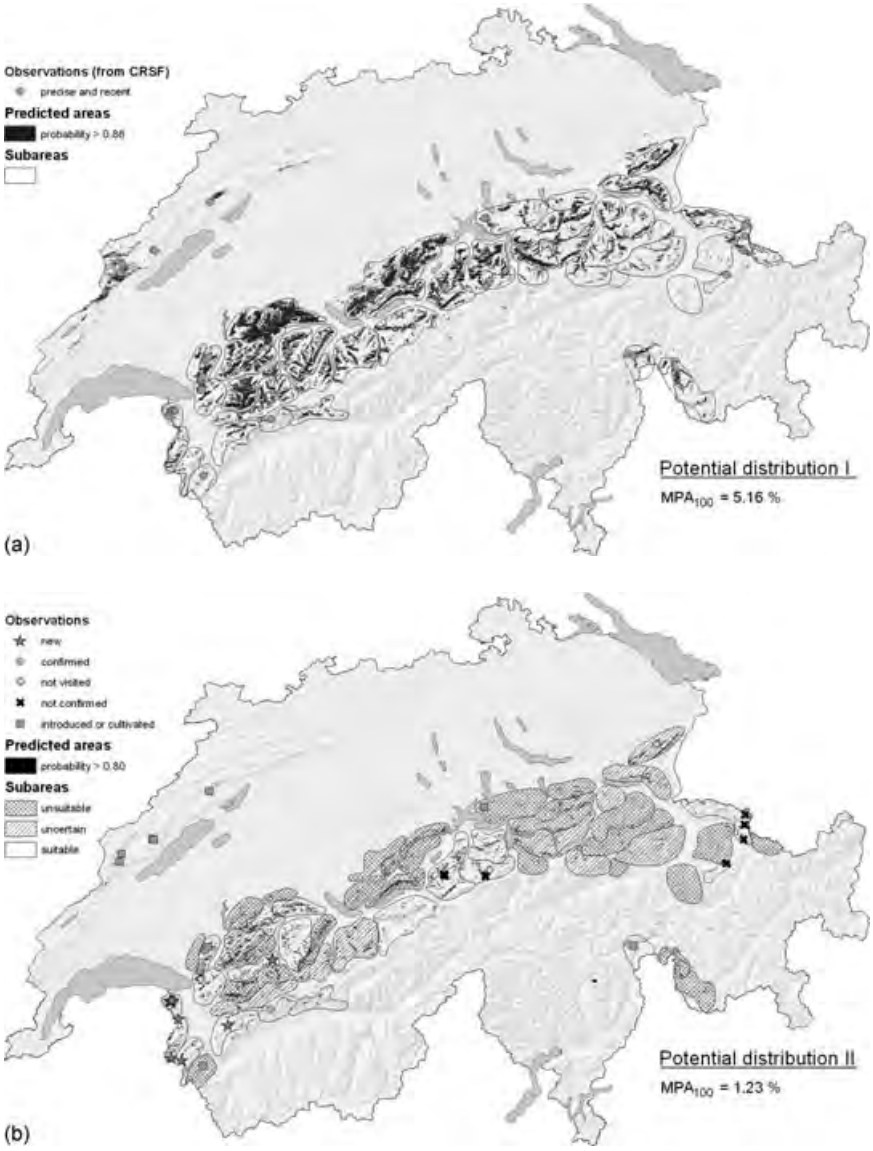


Figure 2. Potential distribution maps for Eryngium alpinum *in Switzerland. (a) Model 1, used to stratify the sampling. Subareas used for sampling are also shown (polygons; see text). Observations correspond to recent (1995 or later) and precise (25-m accuracy) observations. (b) Model 2, improved from model 1 with data sampled in the field. Updated knowledge on geological units important for the species, derived from the field campaign, is also shown (polygons with various shading). The predicted area appears in both maps as dark grey and corresponds to the minimal predicted area (MPA; see methods) containing 100% of observations (MPA_{100}).*

could result if only some geographic areas were subjectively sampled. Because visual detectability was high for *E. alpinum*, absences generated in this way are likely to be reliable. This is unlikely to hold for inconspicuous species; thus field sampling and modeling approaches, including detectability (Stauffer et al. 2002; Edwards et al. 2005; MacKenzie et al. 2005), should be used in these cases. The status of each binary class of qualitative predictor used as a filter was also revised after field sampling (Fig. 1) by checking that no new presence occurred in any of them.

Testing the Approach by Simulation

To test sampling improvement over several iterations, we conducted simulations with a virtual species in a real landscape (Hirzel & Guisan 2002). The distribution of a virtual species can be defined from the available environmental predictors by specifying response curves for each predictor and combining them to draw a habitat suitability map. The final species distribution map is then obtained by cutting the previous suitability map into a presence-absence map reflecting presence and absence of the virtual species in each pixel of the study area. The advantage here is that the truth is known; thus model predictions can always be compared with this original "true" distribution of the species.

In our case, the true distribution of the virtual species in the study area was artificially derived from the first model for *E. alpinum* so that it remained as close as possible to a real situation (same type of distribution, same study area). To turn the probabilistic map into a binary presence-absence map, we used a cut-off value of 0.8, yielding a distribution restricted to about 5% of the study area. This map reflected the true distribution of the virtual species. This way, field iterations could be simulated

simply by checking the presence or absence of the virtual species at each sampling site on this map. After each iteration, the predicted distribution map was reclassified using the MPA threshold—as required by the model-based stratification of the sampling—and the resulting binary predictions were compared with the virtual distribution with the kappa coefficient (Cohen 1960). We used kappa here instead of the AUC because the comparisons involved directly two binary variables. Hence it provided a reliable measure of model improvement over time (here, the successive iterations), reflecting a convergence, or no convergence, toward the true distribution. For computational reasons the simulations were not run over the entire Swiss landscape but on a subset of 500,000 pixels selected in a random-stratified way across the main environmental gradients. This subset of pixels accurately reflected the prevalence of the different habitat suitability classes of the virtual species.

The iterative, model-based, random-stratified sampling procedure was simulated as follows:

1. An initial data set of 30 presences and 30 absences was randomly selected among the pools of 92 presences and 2380 absences used to calibrate the "true distribution" of our virtual species. Hereafter we refer to this data set as the training data set.
2. Using the training data set, a GAM was fitted and predictions were made across the entire study area (i.e., the 500,000 randomly chosen pixels). We used the MPA threshold to transform probabilistic predictions into binary values (presence or absence of the species).
3. The agreement between the binary predictions map and the known binary distribution of the virtual species was evaluated using the kappa coefficient.
4. An equal number of sample sites were selected randomly within (predicted presence) and outside (predicted absence) the MPA (i.e., the two sampling strata). To avoid spatial autocorrelation we did not allow the selection of new sample sites within a radius of 500 m of any already sampled site.
5. The fieldwork was simulated by checking, from the virtual distribution, for true presence or absence in each sample site. At this stage we recorded the number of new occurrence sites found at each iteration. All new and updated observations were then added to the training data set.
6. Steps 2 to 5 were repeated 10 times (e.g., simulating a 10-year survey). Because the selection of the new sample sites involved randomness, we repeated each loop 20 times to quantify the related uncertainty (Fig. 3).

We ran this simulation procedure for three different experiments: sample 30, sample 60, and sample 120. Numbers indicate the initial number of observations and the number of sites sampled at each iteration (e.g., 30 means 15 sites sampled within the "predicted presence" strata and 15 sites within the "predicted absence" strata).

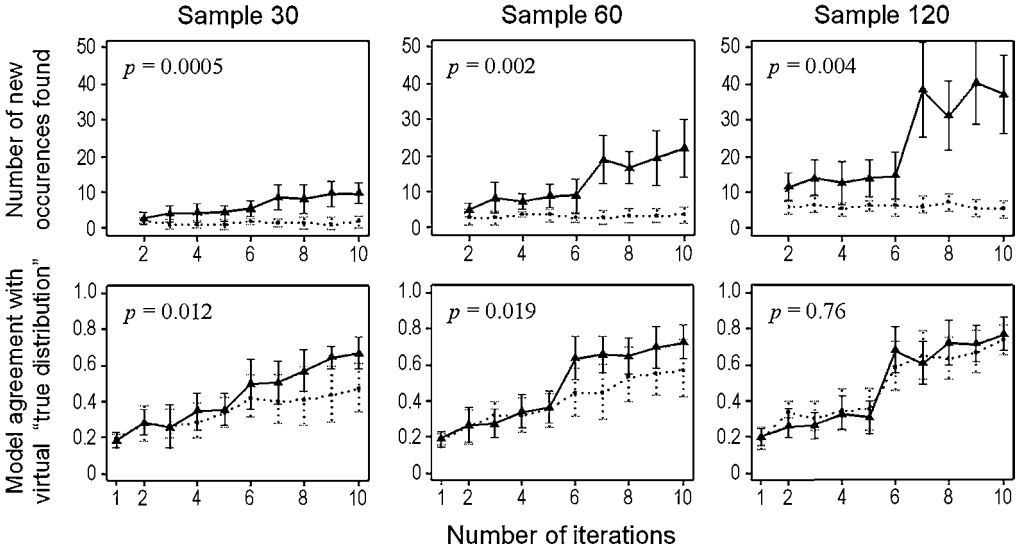


Figure 3. Average (±SE) number of new occurrences of Eryngium alpinum *at each iteration (upper graphs), and the kappa agreement between the predictive model built at each iteration and the true distribution of the virtual species (lower graphs). Solid lines represent simulation with model-based stratification when selecting the new locations to sample. Dashed lines are the control simulation, where the new points to be sampled were chosen completely at random. The horizontal series of graphs correspond to increasing number of sample points visited at each iteration (30, 60, 120). The* p *values are only indicative because they depend on the number of simulations. Each point on a graph was obtained as an average of 20 runs. The field work of iteration* i + 1 *is done using the model of iteration* i*. This explains why the "number of new occurrences found" starts at iteration 2.*

Additionally, control simulations in which the new sample sites were chosen completely at random in the landscape, rather than on a model-based stratification basis, were run for each scenario (Fig. 3). We repeated these control simulations 20 times. Differences between model-based and full-random-sampling simulations were assessed with paired t tests or Wilcoxon signed rank tests when normality was not satisfied. When values were $>$ 0.05, the difference between the two curves was not significant.

Results

A New Picture of the Species' Distribution

Hardly any of the records from the eastern part of the Swiss Alps were confirmed by revisiting existing occurrence localities in 2003, thus offering a very different view of the species range at the end of the field campaign from that suggested by the historical data records. Based on confirmed records only, the status of the species in the Swiss red list would qualify for a downgrade from vulnerable to endangered. Interestingly, most of the unconfirmed records seemed to be specimens that had escaped from nearby gardens, where the plant was cultivated artificially, and in most cases in areas that do not appear naturally suitable to the species long-term survival (when compared with sites where the species was observed).

Model-Based, Random-Stratified Sampling for *E. alpinum*

The first predictive model for *E. alpinum* yielded values of simple evaluation and cross-validation AUC of 0.97 and 0.96, respectively, and the corresponding map was filtered using several qualitative variables (Fig. 2a). Positive filter classes for the species were land use (1, bushes in a 200-m buffer zone from meadow/pasture; 2, rocks in a 200-m buffer zone from meadow/pasture; 3, drain stone; 4, meadow/pasture) and geology (rock and soil) (1, sand and silts; 2, coarse screes [fallen blocks]; 3, marl; 4, marly schist; 5, calcareous phyllites; 6, massive limestone; 7, sandy limestone). For any other qualitative class, predictions were reset to 0. The 100% threshold for this first map was at $p = 0.86$, and the MPA covered 5.1% of the entire Swiss territory. The resulting binary map was superimposed on the shaded relief and divided into subareas (Fig. 2a). A random selection of nearly half the subareas (45.7%, covering 1438 km^2) was visited during summer 2003. Seventy-nine survey walks adequately sampled the two sampling strata, with 56.9% of the paths in areas outside MPA ($p < 0.86$) and the remaining 43.1% of paths within MPA areas ($p \geq 0.86$).

Overall, seven new populations were discovered, which were not previously recorded in the database or the literature. All the new populations occurred in western Switzerland, within highly suitable landscape patches (probability of presence $>$ 91%), and were subsequently used to fit an improved model for the species (Fig. 1).

The second model, improved from a field-based update of the initial data (new observations, update of historical records, and a set of field-checked absences), and revised filters (land use 1 and geology [1–3] removed; classes of slope and aspect added) provided a much narrower prediction of the species range (MPA of 1.2%) that better reflected its actual niche, distribution (local and patchy), and endangered status in the Swiss Alps (Fig. 2b). The same predictors were retained; thus only their coefficients were updated. Most subareas where the species was previously predicted by the primary model, but where no species occurrence could be found, no longer included any suitable patch in the map drawn from the improved model (Fig. 2b). This second model and related spatial predictions had simple validation and cross-validation AUC coefficients of 0.99 and 0.98, respectively.

Simulations

In all three scenarios the number of new presences found over 10 iterations was always much higher when using the model-based stratification approach (about four times in all simulations; Fig. 3). For instance, at a sample size of 30, only 6 iterations were needed with the model-based sampling to double the initial number of presences (i.e., to pass from 30 to 60), whereas nearly 20 iterations would have been needed with simple random sampling. In this case, the model-based approach was 3.3 times faster, corresponding to a net gain in time of 70%. At a sample size of 60, the model-based approach was still 2.3 times faster (considering the doubling case), allowing a 55% gain of time compared with random sampling. In all experiments, the cumulative number of new occurrences found by model-based sampling after 10 iterations was always about 4 times greater than those found by a simple random sampling.

Averages and standard errors of the adequacy between the predictions and the true distribution of the virtual species are also reported for all simulations (Fig. 3). Except for the sample120 scenario, where no significant difference was found, the model-based sampling approach led, on average, to a predicted distribution that had higher agreement with the true distribution of the virtual species than that obtained with the random sampling.

Discussion

We have presented and illustrated the use of predictions obtained from niche-based species distribution models for stratifying sampling and improving data sets on rare and endangered species. Similar approaches based on simple models (e.g., bioclimatic envelope models) have

been published independently (Parris 2002; Poon & Margules 2004). The procedure has the advantage of fitting an existing sampling design in ecology and conservation biology (random-stratified approach), and it adds a new model-based sampling scheme that is expected to increase sampling efficiency, particularly the chance to discover new populations of rare species. The approach was successfully verified through field checking—with seven new populations discovered for the severely endangered species *E. alpinum*—and in computer simulations.

Robust Modeling Approaches Needed for Stratification

To be efficient, the stratification ideally has to be based on as robust an initial model as possible to reduce costs by reiterating only a few sampling stages. Thus the modeling strategy itself is important. Four elements can contribute to a successful modeling effort when species data from such heterogeneous data banks are used. First, species occurrences need to be carefully selected before conducting the statistical analyses by considering the positional accuracy of site location (when recorded in the database). In our study we dropped several occurrences from the original data set, keeping only those corresponding to the grain size and positional accuracy used for the modeling. Such operations limit measurement errors significantly (Engler et al. 2004).

Second, close attention needs to be given to preparing and selecting predictors, including only those expected to have a causal physiological effect on the species (Austin 2002).

Third, using an appropriate modeling technique for the data at hand (e.g., when sufficient occurrences are available, use a flexible semiparametric modeling approach such as GAM) allows for the calculation of response curves that more closely fit the data (Guisan et al. 2002; Lehmann et al. 2002*a*). This is particularly appealing when the predictions are to be made in the same area that was used to calibrate the model. Response plots in GAMs may additionally allow one to check plot density along each environmental predictor and improve the sampling of unevenly sampled predictors during the next field campaign. When fewer data are available, more simple approaches such as climatic envelopes (e.g., BIOCLIM) can be used instead (Guisan & Thuiller 2005). Finally, fourth, when absences are not available from a systematic survey, generating pseudoabsences from presence sites of other rare species seems a promising approach (A.L., unpublished).

Use of Survey Routes

Survey routes recorded on GPS may be a valuable way to explore sampling frameworks further, as a means to lower survey costs and provide additional absences for improving the models. As our results show, generating absences along survey routes can be useful to increase the discriminatory power of the models. Some important limitations exist, however, and should be highlighted before this approach can be further promoted.

Although our species was easily detectable, which was convenient for the first test of the approach, many species will indeed be much harder to detect in the field. Species detectability (MacKenzie et al. 2005) is a major component to consider because it can affect the whole model-based approach in two ways: (1) as additional cost during the field sampling (e.g., more time might be needed to detect cryptic than conspicuous species) and (2) in the way absences are derived from survey routes because these absences might not have the same reliability for different species. The latter aspect is of prime importance when presence-only data are the sole initial source of species data available at the start of the study, as is the case with data from natural history collections (Graham et al. 2004; Rushton et al. 2004).

The use of survey routes across stratifying gradients (in our case, predicted presence or absence of the species) is specific to cases where survey costs need to be reduced. The approach is similar to the cost-reduction "gradsec" sampling technique described by Austin and Heyligers (1989), although applied to a model-based stratified sampling situation rather than to an environmentally based stratification of sampling. Refinements might come here from such guided transects across the stratifying gradients (Ståhl et al. 2000; Thompson 2004) rather than survey routes.

In cases where there is no need to reduce survey costs, a strict model-based, random-stratified approach (i.e., no survey route, no gradsec) in which target cells are randomly chosen in each prediction strata throughout the whole study area should be preferred. The latter is the approach that was used in our simulations.

Confirmation by Simulations

Our simulation results showed that using a model-based rather than a simple random sampling (control) yields both a greater number of presences and a higher adequacy of model predictions. Obtaining such improvement over a random sampling is not surprising because no one would sample rare species randomly in the field, but it provides a standardized and objective measure of improvement. This is, however, mostly true below a certain sample size. With 120 sites (or more) sampled at each iteration, the model-based approach did not converge significantly faster toward the true distribution than the simple random sampling, but the number of new presences found remained much greater (nearly four times). With a sample size of 30 and 60 sites visited at each iteration, the improvement was very apparent, with a significantly greater number of presences found. Thus, the fewer the number of sites sampled at each iteration, the more valuable the use of

the model-based approach. This has direct implications for reducing costs of surveys in nature conservation.

A Flexible, Adaptive Approach

One advantage of this approach is to let the model reflect the current state of a species distribution (or of a set of species). In this sense, it is adaptive. This is particularly important because some environmental conditions or the fitness of some populations might change during the course of the survey, possibly resulting in a temporary lower model evaluation. Such a temporary decrease in model performance may also happen under stable conditions, as observed by simulations (internal variation within the 20 replicates; see error bars in Fig. 3). In the latter cases, the model returned rapidly (usually within the next iteration) to a situation of higher agreement with the true geographic distribution. The same might be expected after an environmental change, which is thus only likely to delay model convergence by some number of iterations.

Which Method for Which Data?

In this study, we had recourse to GAMs to predict the distribution of a rare species for which 92 occurrences were initially available. GAMs, however, are relatively data-hungry modeling techniques. As such they need a sufficient number of occurrences to be fitted. In many cases fewer occurrences will be available at the start of the study, as in the case of many rare and endangered species in data-poor countries, which will prevent the use of such an advanced technique (and others of similar complexity). When too few species occurrences are available (say < 20), alternative approaches must be found. Alternatives can be (1) a simpler technique such as climatic envelopes (Poon & Margules 2004); (2) a modeling approach at the community level (Ferrier et al. 2002), where the data for more common species may help support the modeling of less frequent species; or (3) models for more common species that are frequently associated with the rare target species. In the last case, Edwards et al. (2005) show that a model-based stratified sampling based on common species can significantly improve detectability of rare species in coarse-scale surveys.

The resolution used in our study was also much finer than often available elsewhere. Although this approach could theoretically be conducted at any resolution, problems may well appear when the size of the sampling unit becomes too large to be properly surveyed in the field. Again, the choice of an appropriate resolution and survey design is likely to depend primarily on a species' detectability (screening a large sampling cell is easier for a conspicuous than for a cryptic species). Future tests of the method may thus need to take detection probability into account in the sampling design and related statistical inferences (MacKenzie et al. 2005).

Sampling Bias and Predictors Used

There might be a risk of converging toward a biased model, because of the influence of some of the initial parameters (e.g., initial sample size, number of sites sampled at each field iteration, initial environmental or geographic bias in the data). As long as the area is sampled probabilistically in both the suitable and unsuitable strata, however, the model iterations should converge rapidly toward unbiased estimates.

Besides an initial sampling bias, predictions may additionally be weakened—and thus with it the strength of the approach—if one or several proximal environmental predictors (Austin 2002) are not included in the initial model or if the model is severely overfitted. It is thus critical to decide which predictors should be included and how strict the limitation of the final number of predictors in the model should be. With only a few species occurrences, a limitation rule that is too strict (e.g., a stepwise procedure in a regression model, based on the Bayesian information criterion) may prevent the inclusion of some proximal predictors, whereas including too many predictors may cause the model to direct the sampling too narrowly toward some situations. In both cases the effect is likely to be a reduction in sampling efficiency. Hence a tradeoff has to be found, which still requires further investigation.

Future Research Directions

Where multiple species and overall biodiversity are the focus, distinct sampling designs could be used (Ferrier 2002; Edwards et al. 2005) or the same model-based, random-stratified design might be adapted to a multispecies situation. The use of alternative modeling methods such as community modeling (Ferrier 2002; Ferrier et al. 2002) should also be investigated, within a similar model-based sampling framework, when too few occurrences are available at the start of the study but more common species were jointly inventoried (which was not the situation in our case).

Improved field design may also be implemented within such a model-based approach, such as adaptive cluster sampling (Thompson & Seber 1996). In adaptive cluster sampling, a more intensive search of the species in neighbor plots is made every time the species is met at one of the plots included in the original design. The search is conducted within a neighborhood window of specified shape. Every time the species is found again within this window, the same procedure is repeated until no additional occurrence is found. This design is optimal when improved estimates of the whole population size in the area are requested and is particularly efficient for sampling rare, clustered species (Thompson & Seber 1996; Christman 2004). Hence it will be particularly suited for the case of rare species with large populations in the areas where they occur (i.e., the "urban" type described in

Collins et al. [1993]). Such adaptive cluster sampling was for instance successfully tested in a parallel study of an invasive species—*Heracleum mantegazzianum*—that has still a scattered distribution in the study area (A.G., unpublished results).

Model-based sampling of rare species, involving reiterative alternation of modeling and field sampling phases, shows great promise for strengthening and complementing conservation practices and reducing sampling costs. It would be optimal if applied together with the monitoring of those additional parameters required by population management (e.g., accessibility of sites as a surrogate for picking; agricultural practices as a surrogate for grazing). The approach would still deserve more thorough testing, however, especially in the field, before it can be applied more widely. Once more broadly validated, it may be seen as a valuable approach for nature conservation agencies.

Acknowledgments

We thank B. Baümler at the Swiss Floristical Center in Geneva for making the species data available; L. Rechsteiner, C. Trippi, and Y. Naciri-Graven for additional species data or for help with the field work; and M. Burgman and two anonymous reviewers for useful comments on an earlier version of the manuscript. We further thank the MAVA Foundation (Switzerland), the European Union through the FP6 project EVK3-2003-505373-IntraBioDiv, and the Biodiversity Programme of the Norwegian Research Council for financial support.

Literature Cited

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., and P. C. Heyligers. 1989. Vegetation survey design for conservation: gradsec sampling of forests in north-east New South Wales. Biological Conservation **50:** 13–32.

Brook, B. W., J. J. O'Grady, A. P. Chapman, M. A. Burgman, H. R. Akcakaya, and R. Frankham. 2000. Predictive accuracy of population viability analysis in conservation biology. Nature **404:** 385–387.

Christman, M. C. 2004. Sequential sampling for rare and geographically clustered populations. Pages 134–145 in W. L. Thompson, editor. Sampling rare or elusive species. Island Press, Washington, D.C.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement **20:** 37–46.

Collins, S. L., S. M. Glenn, and D. W. Roberts. 1993. The hierarchical continuum concept. Journal of Vegetation Science **4:** 149–156.

Côté, I. M., and J. D. Reynolds. 2002. Predictive ecology to the rescue? Science **298:** 1181–1182.

Edwards, T. C. J., R. Cutler, N. E. Zimmermann, L. Geiser, and J. Alegria. 2005. Model-based stratifications for enhancing the detection of rare ecological events: lichens as a case study. Ecology **86:** 1081–1090.

Elith, J., and M. A. Burgman. 2002. Predictions and their validation: rare plants in the central highlands, Victoria, Australia. Pages 303–313 in J. M. Scott, P. J. Heglund, J. B. Haufler, M. Morrison, M. G. Raphael, W. B. Wall, and F. Samson, editors. Predicting species occurrences: issues of accuracy and scale. Island Press, Covelo, California.

Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology **41:** 263–274.

Ferrier, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? Systematic Biology **51:** 331–363.

Ferrier, S., M. Drielsma, G. Manion, and G. Watson. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. Biodiversity and Conservation **11:** 2309–2338.

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.

Graham, C. H., S. Ferrier, F. Huettman, C. Moritz, and A. T. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution **19:** 497–503.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters **8:** 993–1009.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling **135:** 147–186.

Guisan, A., T. C. Edwards, and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling **157:** 89–100.

Hastie, T., and R. Tibshirani. 1986. Generalized additive models. Statistical Sciences **1:** 297–318.

Hirzel, A., and A. Guisan. 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling **157:** 331–341.

IUCN (World Conservation Union)/SSC (Species Survival Commission). 2001. IUCN Red List categories. IUCN, Gland, Switzerland.

Lamoreux, J., et al. 2003. Value of the IUCN Red List. Trends in Ecology & Evolution **18:** 214–215.

Lehmann, A., J. M. Overton, and M. P. Austin. 2002*a*. Regression models for spatial prediction: their role for biodiversity and conservation. Biodiversity and Conservation **11:** 2085–2092.

Lehmann, A., J. M. Overton, and J. R. Leathwick. 2002*b*. GRASP: generalized regression analysis and spatial prediction. Ecological Modelling **157:** 189–207.

MacKenzie, D. I., J. D. Nichols, N. Sutton, K. Kawanishi, and L. L. Bailey. 2005. Improving inferences in population studies of rare species that are detected imperfectly. Ecology **86:** 1101–1113.

Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology **38:** 921–931.

Miller, R. I. 1986. Predicting rare plant distribution patterns in the southern Appalachians of the south-eastern U.S.A. Journal of Biogeography **13:** 293–311.

Parris, K. M. 2002. The distribution and habitat requirements of the great barred frog (*Mixophyes fasciolatus*). Wildlife Research **29:** 469–474.

Poon, E. L., and C. R. Margules. 2004. Searching for new populations of rare plant species in remote locations. Pages 189–207 in W. L. Thompson, editor. Sampling rare or elusive species. Island Press, Washington, D.C.

Prendergast, J. R., R. M. Quinn, J. H. Lawton, B. C. Eversham, and D. W. Gibbons. 1993. Rare species, the coincidence of diversity hotspots and conservation strategies. Nature **365:** 335–337.

Rushton, S. P., S. J. Ormerod, and G. Kerby. 2004. New paradigms for modelling species distributions? Journal of Applied Ecology **41:** 193–200.

Ståhl, G., A. Ringvall, and T. Lamas. 2000. Guided transect sampling for assessing sparse populations. Forest Science **46:** 108–115.

Stauffer, H. B., C. J. Ralph, and S. L. Miller. 2002. Incorporating detection uncertainty into presence-absence surveys for Marblet Murrelet.

Pages 357–365 in J. M. Scott, P. J. Heglund, M. L. Morrison, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: issues of accuracy and scale. Island Press, Covelo, California.

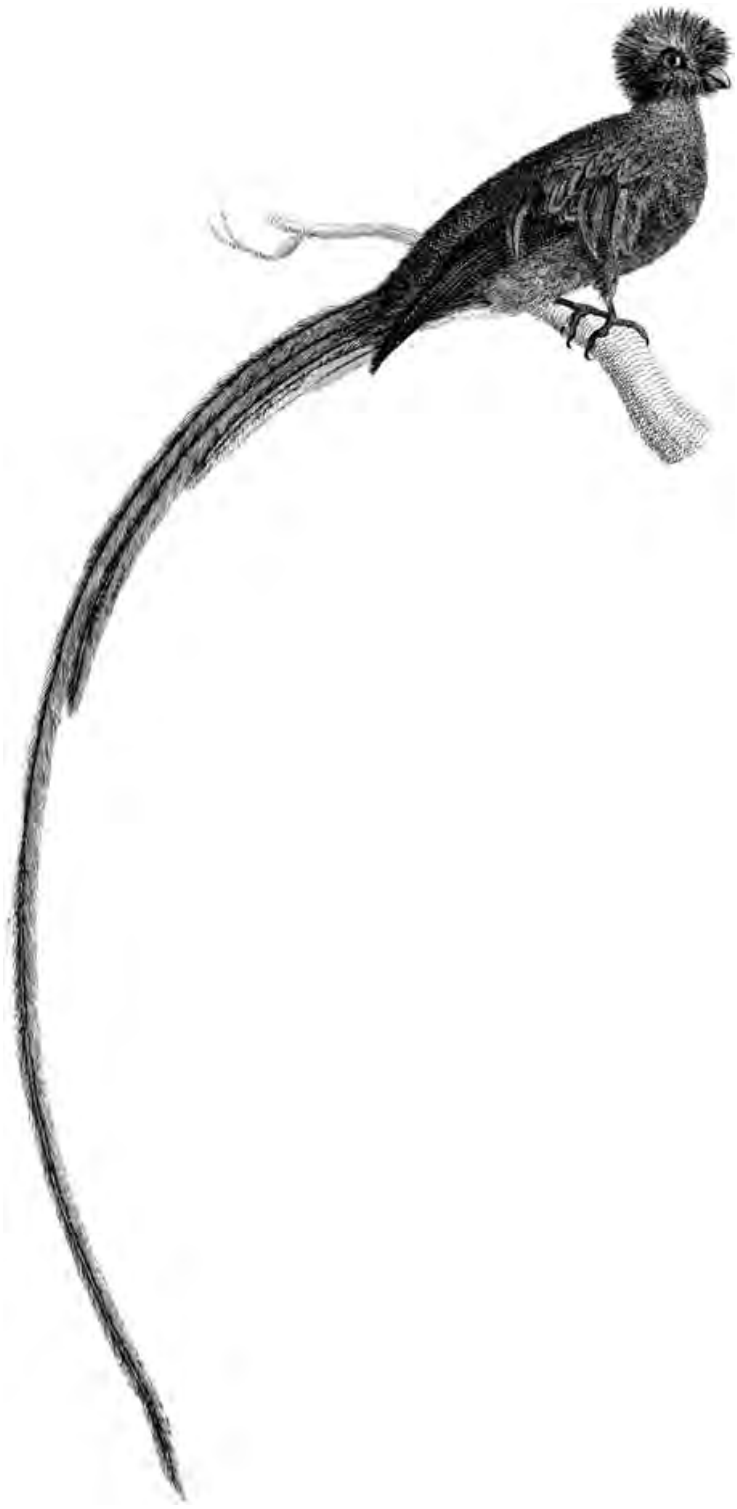
Thompson, S. K., and G. A. Seber. 1996. Adaptive sampling. Wiley, New York.

Thompson, W. L. 2004. Future directions in estimating abundance of rare or elusive species. Pages 389–399 in W. L. Thompson, editor. Sampling rare or elusive species. Island Press, Washington, D.C.

Vaughan, I. P., and S. J. Ormerod. 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. Conservation Biology **17:** 1601–1611.

Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. Trends in Ecology & Evolution **16:** 446–453.

Zimmermann, N. E., and F. Kienast. 1999. Predictive mapping of alpine grasslands in Switzerland: species versus community approach. Journal of Vegetation Science **10:** 469–482.



A 2

BIOMOD - a platform for ensemble forecasting of species distributions.

Thuiller W., Lafourcade B., Engler R. and Araújo M.B., 2009. *Ecography*, **32**, 369-373.

00000100000
00001010000
00010001000
00100000100
01000000010
10000000001
Software notes

Ecography 32: 369–373, 2009
doi: 10.1111/j.1600-0587.2008.05742.x

Subject Editor: Carsten Rahbek. Accepted 30 December 2008

BIOMOD – a platform for ensemble forecasting of species distributions

Wilfried Thuiller, Bruno Lafourcade, Robin Engler and Miguel B. Araújo

W. Thuiller (wilfried.thuiller@ujf-grenoble.fr) and B. Lafourcade, Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Univ. Joseph Fourier, BP 53, FR–38041 Grenoble Cedex 9, France. – R. Engler, Dépt d'Ecologie et d'Evolution (DEE), Univ. de Lausanne, Bâtiment de Biologie, CH–1015 Lausanne, Switzerland. – M. B. Araújo, Dept of Biodiversity and Evolutionary Biology, National Museum of Natural Science, CSIC, C/Gutierrez Abascal, 2, ES–28006, Madrid, Spain and Rui Nabeiro Biodiversity Chair, Univ. of Évora, Colégio Luís António, Rua Romão Ramalho no. 59, PT-7000-671, Évora, Portugal.

BIOMOD is a computer platform for ensemble forecasting of species distributions, enabling the treatment of a range of methodological uncertainties in models and the examination of species-environment relationships. BIOMOD includes the ability to model species distributions with several techniques, test models with a wide range of approaches, project species distributions into different environmental conditions (e.g. climate or land use change scenarios) and dispersal functions. It allows assessing species temporal turnover, plot species response curves, and test the strength of species interactions with predictor variables. BIOMOD is implemented in R and is a freeware, open source, package.

Species distribution models (SDM, Guisan and Thuiller 2005) are being used in nearly all branches of life and environmental sciences. A quick search in ISI Web of Science (18/02/08) using ''species distribution models'' OR ''niche models'' OR ''habitat models'' OR ''bioclimatic models'' highlights 21 973 papers, 74% of which published in the past 10 yr, in fields as varied as environmental sciences (53% of the records), zoology (15%), marine and freshwater biology (15%), life sciences and biomedicine (9%), biodiversity and conservation (8%), evolutionary biology (8%), fisheries (6%), forestry (6%), oceanography (5%), genetics and heredity (5%), amongst others. Advancement of knowledge in these fields is now intertwined with technical innovation in species distribution modelling and dependent on the existence of suitable software for fitting models and examining results. One difficulty with the use of species distribution models is that the number of techniques available is large and is increasing steadily, making it difficult for ''non-aficionados'' to select the most appropriate methodology for their needs (Elith et al. 2006, Heikkinen et al. 2006). Recent analyses have also demonstrated that discrepancies between different techniques can be very large, making the choice of the appropriate model even more difficult. This is particularly true when models are used to project distributions of species into independent situations, which is the example of projections of species distributions under future climate change scenarios (Thuiller 2004, Pearson et al. 2006). A possible solution to account for this inter-model variability is to fit ensembles of forecasts by simulating across more than one set of initial conditions, model classes, model parameters, and boundary conditions (for a review see Araújo and New 2007) and analyse the resulting range of uncertainties with bounding box, consensus and probabilistic methodologies rather than lining up with a single modelling outcome (Araújo and New 2007, Thuiller 2007). BIOMOD offers such a platform for ensemble forecasting (Fig. 1) using freeware and open-source R software (R Development Core Team 2008). It overcomes some of the limitations of existing software (e.g. being able to fit and compare different models) and incorporates several features for testing models (e.g. k-fold cross validation) and for examining species-environment relationships (e.g. using randomization tests) (Fig. 2).

Earlier implementations of BIOMOD (Thuiller 2003, 2004) provided limited ensemble simulations across model classes (i.e. four modelling techniques) and boundary conditions (i.e. up to five climate scenarios). Currently, BIOMOD enables larger simulations across initial conditions (i.e. by randomly re-sampling species distribution data and fitting different models for each sample), nine model classes (generalised linear models (GLM, McCullagh and Nelder 1989), generalised additive models (GAM, Hastie and Tibshirani 1990), multivariate adaptive regression splines (MARS, Friedman 1991), classification tree analysis (CTA, Breiman et al. 1984), mixture discriminant analysis (MDA, Hastie et al. 1994), artificial neural networks (ANN, Ripley 1996), generalised boosted models (GBM, Ridgeway 1999), random forests (Breiman 2001), and one rectilinear envelope similar to BIOCLIM (SRE, Busby

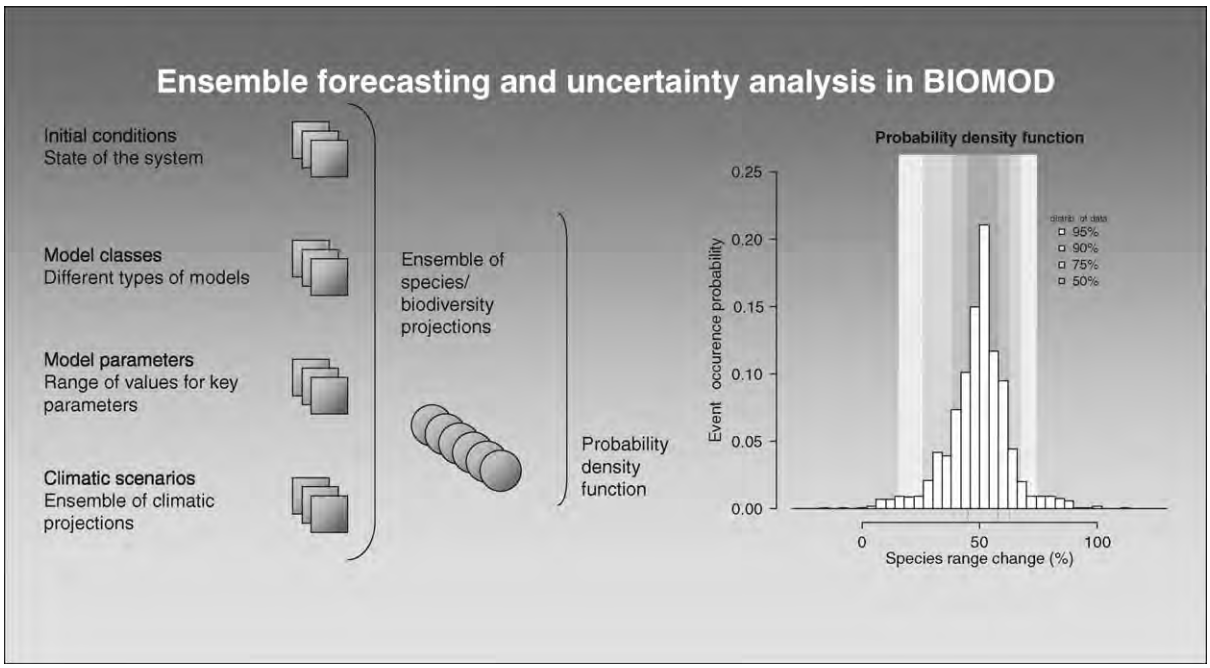


Figure 1. Probabilistic approach for forecasting species potential distributions (adapted from Thuiller 2007).

1991)), a variable number of model parameterizations (e.g. polynomials and smoothing splines of different orders in general linear or additive models, nodes in classification trees, hidden layers in neural nets), and a virtually unlimited number of boundary conditions. Most modelling techniques implemented in BIOMOD require that species distribution data are presence and absence. When data are presence-only, a simple solution is to generate random pseudo-absences. This can be done in BIOMOD using strategies of increasing complexity.

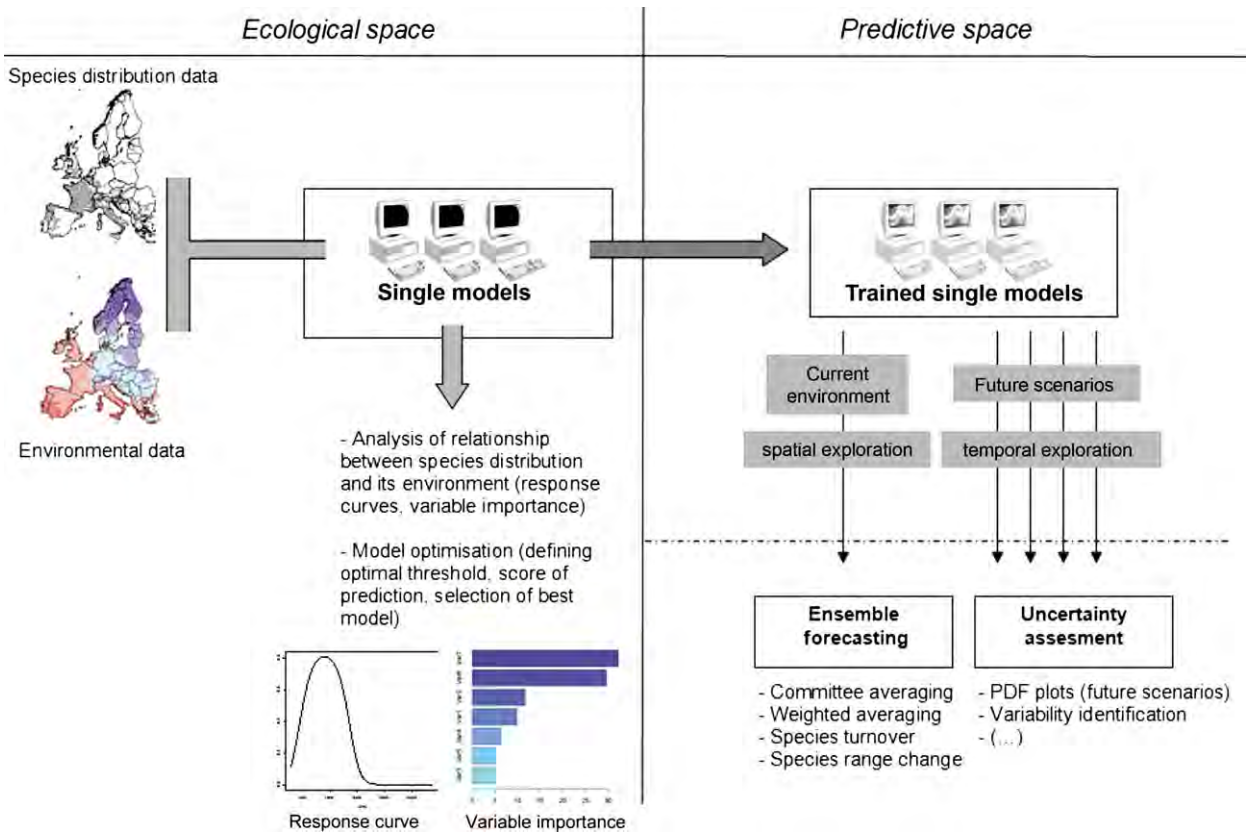


Figure 2. Schematic representation of the modelling procedure in BIOMOD.

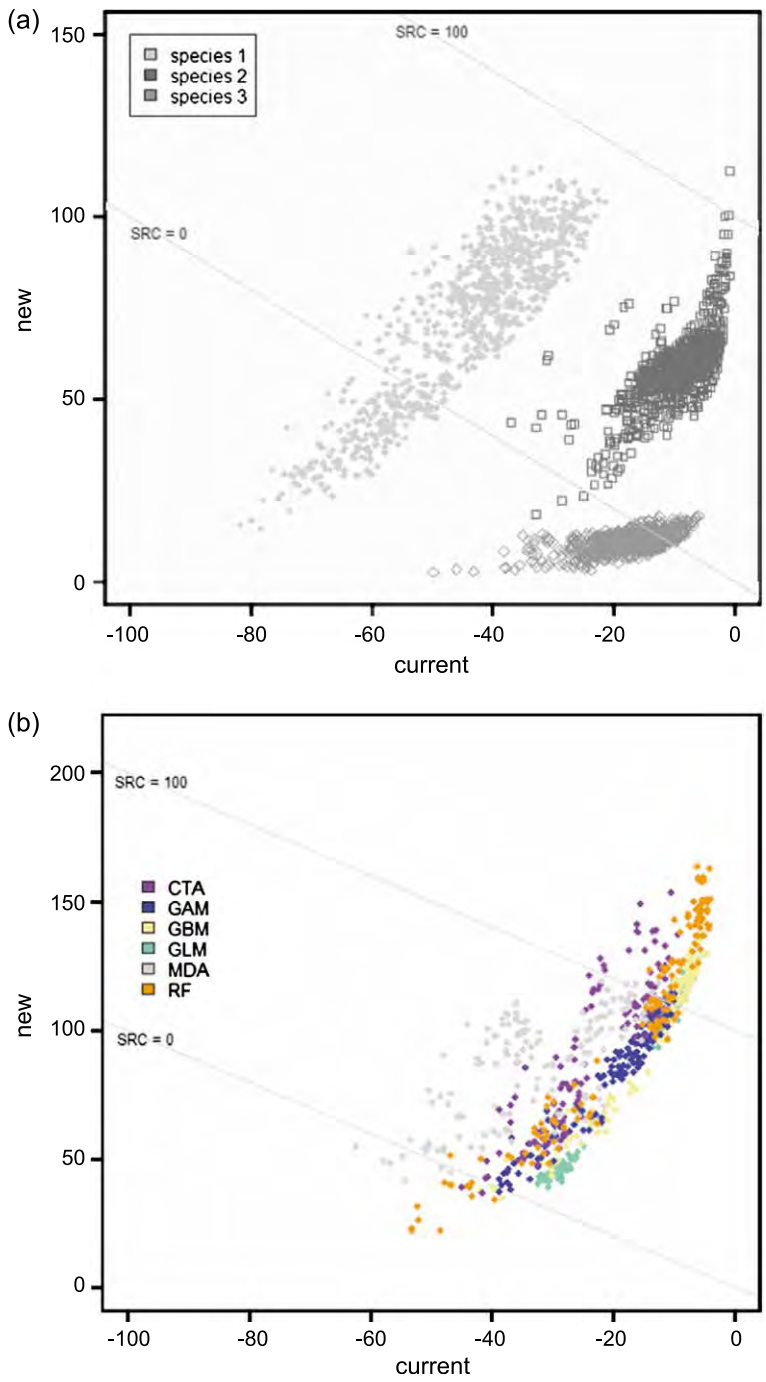


Figure 3. A representation of species potential range changes. Species range changes (SRC) are represented by a 2 axes plot: the first axis (x) represents the percentage of currently occupied sites projected to be lost; the second axis (y) represents the relative percentage – according to the current distribution's size – of currently unsuitable sites but projected to be suitable in future. For each dot, the sum of these two values gives the SRC. Every dot is a projection (e.g. different climate change scenarios, different model parameterisations). (a) Species range change projections for 3 species. Whilst species 2 keeps the majority of its current sites and gains around half of its potential habitat surface, species 3 is projected to keep as many of its sites, but will gain very few new suitable sites. Species 1 will lose suitability in many current sites but will also gain suitability in many new sites. (b) Species range change (SRC) according to different modellin techniques. This type of plot enables visual exploration of the sources of uncertainty accrued from different methodological sources of uncertainty, such as model algorithms, criteria to transform probabilities into presence and absence, climate change scenarios.

Evaluation of models in BIOMOD includes two sorts of analysis: assessments of the goodness-of-fit (=explanatory power) and of model accuracy (=predictive power). The former uses standard approaches associated with each technique; for example, ANOVA decomposition and AIC are available for both GLM and GAM, whereas rate of misclassification is used for CTA. The latter can be performed with three different procedures: the area under the relative operating characteristic curve (AUC, Hanley and McNeil 1982), Cohen's K (Monserud and Leemans

1992), and the true skill statistic (TSS, Allouche et al. 2006).

In an ideal world, model accuracy (e.g. AUC/Kappa/TSS) should always be evaluated with statistically independent data, i.e. training data that are not spatially autocorrelated with test data (Araújo et al. 2005a). When independent data are not available, an alternative is to use data-splitting procedures, whereby a proportion of the original data are used for training the models and the withheld data are used for model evaluation. A single random splitting of data was available in earlier implementations of BIOMOD (Thuiller 2003), but it proved to be a non-negligible source of variability when making predictions. Currently, BIOMOD allows much greater flexibility. Apart from the ability to define the size of the training and test datasets, BIOMOD also allows k number of data splitting runs to be computed. In each run, a model is fitted on one part of the data and evaluated on the left-out data. The evaluation values provided by each of the k splitting runs are then averaged, ensuring the final evaluation is quasi-independent of a particular realisation of random split. BIOMOD also provides a version of ''leave-one-out'' resampling. Users simply need to define the training sets as 100% of the data minus 1 record and then repeat the procedure a user-defined number of times (e.g. 1000 times). When non-independent data are used for model evaluation, variability in model accuracy should be interpreted as a measure of the sensitivity of model results to the initial conditions rather than as a measure of predictive accuracy (Araújo and Guisan 2006).

Model evaluation can be used to investigate the variability of predictions across modelling techniques. In BIOMOD a table displaying the AUC/K/TSS values is produced for each model and for each species. This table can be used for selecting the ''best'' model, i.e. the model providing greater accuracy on the test data for each species (Thuiller 2003). Assuming that no modelling procedure is always better, selecting the best model for each situation might be a useful option. The alternative ensemble forecasting paradigm draws on the assumption that model accuracy on non-independent test data is not representative of model accuracy on independent situations. In such cases, committee averaging of model predictions (giving the same weight to all predictions) can be implemented to derive a consensus prediction; an alternative is to combine models using some form of weighting (e.g. using PCA score value, Thuiller 2004, Araújo et al. 2006). There are a range of approaches to do this (for review see Araújo and New 2007), but in BIOMOD weights are currently calculated on the basis of models' predictive accuracy on test data (i.e. a form of ''stacking''). Empirical testing of consensus forecasting under climate change has shown that weighted approaches are promising (Araújo et al. 2005b, Marmion et al. 2008).

When using models to predict potential distributions in other regions, or times, it is often useful to visually examine species response curves (Austin and Gaywood 1994). To do so, BIOMOD uses an implementation of the ''evaluation strip'' procedure (Elith et al. 2005), making it possible to extract species' response curves independently of the model's algorithm.

There are some techniques available for characterising variable contribution in model predictions. However, these techniques are model-specific, so are the conclusions that one may extract from them. To overcome this limitation, BIOMOD uses a randomisation procedure to estimate the importance of each variable. The procedure is independent of the modelling technique, thus enabling direct comparison across models. This procedure uses Pearson correlation between the standard predictions (i.e. fitted values) and predictions where the variable under investigation has been randomly permutated. If the correlation is high, i.e. it is showing little difference between the two predictions, the variable permutated is considered not important for the model. This is repeated a user-defined number of times for each variable, and the mean correlation coefficient over the runs is kept. BIOMOD then gives a ranking of the variables for each of the model selected.

Finally, when projecting potential distributions of species into future environmental conditions, different dispersal assumptions can be made: no dispersal; unlimited dispersal; and user-defined species-specific dispersal. Measures of temporal turnover in potential species richness can then be calculated for each period (Thuiller et al. 2005), as well as species habitat change (Fig. 3a), and visualized according to the different models used (Fig. 3b) to emphasize the potential uncertainty coming from modelling technique or climate change scenarios' choice.

The BIOMOD R-package and a detailed user's guide of BIOMOD is available at the R-Forge website <biomod.r-forge.r-project.org>. To cite BIOMOD, or acknowledge its use, cite this Software Note as follows, substituting the version of the application that you used for ''Version 0'':

Thuiller, W., Lafourcade, B., Engler, R. and Araújo, M. B. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. – Ecography 32: 369–373 (Version 0).

Acknowledgements – BIOMOD is currently developed with funding from FP6 EU MACIS (No. 044399 SSPI) and ECOCHANGE (GOCE-CT-2003-506675) projects; MBA is also funded by FBBVA BIOIMPACTO project. We are indebted to the R community for their valuable contribution to the development and improvement of R code and libraries.

References

Allouche, O. et al. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). – J. Appl. Ecol. 43: 1223–1232.

Araújo, M. B. and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. – J. Biogeogr. 33: 1677–1688.

Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – Trends Ecol. Evol. 22: 42–47.

Araújo, M. B. et al. 2005a. Validation of species-climate impact models under climate change. – Global Change Biol. 11: 1504–1513.

Araújo, M. B. et al. 2005b. Reducing uncertainty in projections of extinction risk from climate change. – Global Ecol. Biogeogr. 14: 529–538.

Araújo, M. B. et al. 2006. Climate warming and the decline of amphibians and reptiles in Europe. – J. Biogeogr. 33: 1712–1728.

Austin, M. P. and Gaywood, M. J. 1994. Current problems of environmental gradients and species response curves in relation to continuum theory. – J. Veg. Sci. 5: 473–482.

Breiman, L. 2001. Random forests. – Mach. Learn. 45: 5–32.
Breiman, L. et al. 1984. Classification and regression trees. – Chapman and Hall.
Busby, J. R. 1991. BIOCLIM – a bioclimate analysis and prediction system. – In: Margules, C. R. and Austin, M. P. (eds), Nature conservation: cost effective biological surveys and data analysis. CSIRO, pp. 64–68.
Elith, J. et al. 2005. The evaluation strip: a new and robust method for plotting predicted responses from species distribution models. – Ecol. Model. 186: 280–289.
Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – Ecography 29: 129–151.
Friedman, J. 1991. Multivariate adaptive regression splines. – Ann. Stat. 19: 1–141.
Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.
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. et al. 1994. Flexible discriminant analysis by optimal scoring. – J. Am. Stat. Assoc. 89: 1255–1270.
Hastie, T. J. and Tibshirani, R. 1990. Generalized additive models. – Chapman and Hall.
Heikkinen, R. K. et al. 2006. Methods and uncertainties in bioclimatic envelope modeling under climate change. – Prog. Phys. Geogr. 30: 751–777.
Marmion, M. et al. 2008. Evaluation of consensus methods in predictive species distribution modelling. – Divers. Distrib. 15: 59–69.
McCullagh, P. and Nelder, J. A. 1989. Generalized linear models. – Chapman and Hall.
Monserud, R. A. and Leemans, R. 1992. Comparing global vegetation maps with the Kappa statistic. – Ecol. Model. 62: 275–293.
Pearson, R. G. et al. 2006. Model-based uncertainty in species' range prediction. – J. Biogeogr. 33: 1704–1711.
R Development Core Team 2008. R: a language and environment for statistical computing. – R Foundation for Statistical Computing.
Ridgeway, G. 1999. The state of boosting. – Comput. Sci. Stat. 31: 172–181.
Ripley, B. D. 1996. Pattern recognition and neural networks. – Cambridge Univ. Press.
Thuiller, W. 2003. BIOMOD: optimising predictions of species distributions and projecting potential future shifts under global change. – Global Change Biol. 9: 1353–1362.
Thuiller, W. 2004. Patterns and uncertainties of species' range shifts under climate change. – Global Change Biol. 10: 2020–2027.
Thuiller, W. 2007. Biodiversity – climate change and the ecologist. – Nature 448: 550–552.
Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. – Proc. Nat. Acad. Sci. USA 102: 8245–8250.

A 3

Climate change and plant distribution: local models predict high-elevation persistence.

Randin C.F., Engler R., Normand S., Zappa M., Zimmermann N.E., Pearman P.B., Vittoz P., Thuiller W., Guisan A., 2009. Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology,* **15**, 1557-1569.

Global Change Biology (2009) **15**, 1557–1569, doi: 10.1111/j.1365-2486.2008.01766.x

Climate change and plant distribution: local models predict high-elevation persistence

CHRISTOPHE F. RANDIN*, ROBIN ENGLER*, SIGNE NORMAND†, MASSIMILIANO ZAPPA‡, NIKLAUS E. ZIMMERMANN‡, PETER B. PEARMAN*, PASCAL VITTOZ§, WILFRIED THUILLER¶ and ANTOINE GUISAN*

Laboratory for Conservation Biology, Department of Ecology and Evolution, Biophore, CH-1015 Lausanne, Switzerland, †Department of Biological Sciences, University of Aarhus, Ny Munkegade, Building 1540, DK-8000 Aarhus C, Denmark, ‡Swiss Federal Research Institute WSL, Land Use Dynamics, Zuercherstrasse 111, CH-8903 Birmensdorf, Switzerland, §Faculté des Géosciences et de l'Environnement (GSE), University of Lausanne, CH-1015 Lausanne, Switzerland, ¶Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France

Abstract

Mountain ecosystems will likely be affected by global warming during the 21st century, with substantial biodiversity loss predicted by species distribution models (SDMs). Depending on the geographic extent, elevation range, and spatial resolution of data used in making these models, different rates of habitat loss have been predicted, with associated risk of species extinction. Few coordinated across-scale comparisons have been made using data of different resolutions and geographic extents. Here, we assess whether climate change-induced habitat losses predicted at the European scale ($10 \times 10'$ grid cells) are also predicted from local-scale data and modeling (25 m × 25 m grid cells) in two regions of the Swiss Alps. We show that local-scale models predict persistence of suitable habitats in up to 100% of species that were predicted by a European-scale model to lose all their suitable habitats in the area. Proportion of habitat loss depends on climate change scenario and study area. We find good agreement between the mismatch in predictions between scales and the fine-grain elevation range within $10 \times 10'$ cells. The greatest prediction discrepancy for alpine species occurs in the area with the largest nival zone. Our results suggest elevation range as the main driver for the observed prediction discrepancies. Local-scale projections may better reflect the possibility for species to track their climatic requirement toward higher elevations.

Keywords: climate change, Europe, mountain region, species distribution model, Swiss Alps

Received 19 June 2008 and accepted 8 September 2008

Introduction

Mountain ecosystems are likely sensitive to global warming owing to the reduction in area with increasing elevation (Guisan *et al*., 1995; Theurillat *et al*., 1998; Diaz *et al*., 2003; Beniston, 2006). A recent global assessment of the impacts of climate change on these ecosystems suggests that they should experience unprecedented rates of warming during the 21st century, two to three times greater than observed during the 20th century (Nogués-Bravo *et al*., 2006). These rapid changes in temperature and other climate parameters at high elevations are expected to have strong effects on plant communities (Guisan *et al*., 1995; Beniston *et al*., 1996; Guisan & Theurillat, 2000; Walther, 2003). The first biological impacts of past and ongoing global warming are already visible in the Alps, and include the upward shift of treelines (Gehrig-Fasel, 2007) and the upward shift and range reduction in alpine and nival plant species (Braun-Blanquet, 1957; Hofer, 1992; Grabherr *et al*., 1994; Pauli *et al*., 1996, 2007; Walther *et al*., 2005; Vittoz *et al*., 2006).

In the last decade, species distribution models (SDMs; Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005) have become important tools to evaluate the potential impacts of climate change on plant distributions (Bakkenes *et al*., 2002; Thomas *et al*., 2004; Thuiller *et al*., 2005). These tools statistically relate multiple abiotic habitat characteristics (*sensu* Kearney & Porter, 2004)

Correspondence: Antoine Guisan, tel. + 41 21 692 42 54, fax + 41 21 692 42 65, e-mail: antoine.guisan@unil.ch

with observed occurrences of a species, thus fitting the original definition of the Hutchinsonian (Hutchinson, 1957) environmental niche (Kearney & Porter, 2004; i.e. without explicitly requiring a mechanistic link between environmental gradients and population fitness (see also Guisan & Thuiller, 2005; Araujo & Guisan, 2006). Hereafter, we will simply refer to the *realized niche* for this initial definition of the niche.

Using these tools, models of climate change impacts on biodiversity have been developed at several scales (but see Davis *et al*., 1998; Bahn & Körner, 2003; Hampe, 2004; Dormann, 2007). At the extent of western Europe, Thuiller *et al*. (2005) forecasted that the plant diversity of some European mountain ranges (e.g. mid-elevation Alps) could be disproportionably sensitive to climate change, with up to 60% species loss per 10′ grid cell. This study modeled the distribution of 1350 species using data from the Atlas Florae Europaeae (AFE; Lahti & Lampinen, 1999). Similarly, a local-scale study of 85 subalpine and alpine nonwoody plants of open habitat was conducted at high resolution (20 m × 20 m) in the Austrian Alps (Dirnböck *et al*., 2003). The authors predicted that up to 40–50% of the plant species could potentially become extinct owing to climate change. Finally, in a study of 62 alpine and nival plants, Guisan & Theurillat (2000) predicted relatively low rates of total habitat loss, between 2% and 5%, but nearly 40% of the species were nevertheless predicted to lose more than 90% of their suitable habitat. These SDM results suggest that alpine and nival plants, in particular, may lose much of their suitable climatic habitats, owing to the decrease in suitable habitat area with increasing elevation (Guisan & Theurillat, 2000). Nonetheless, still few studies have used SDMs to assess the possible impacts of climate change on plant species in mountain environments (e.g. Dirnböck *et al*., 2003).

Although SDM studies consistently predict substantial impacts on plant diversity in mountains, the rates of predicted habitat loss vary among studies, which themselves vary in study area extent, data resolution, and species composition. In spite of this and of variation in the results of the studies, very few comparisons of predictions across resolutions in a single region have been attempted (e.g. Trivedi *et al*., 2008). This is potentially an important deficiency because coarse-resolution predictions based on SDMs are commonly used in the preparation of reports by the Intergovernmental Panel on Climate Change (IPCC; http://www.ipcc.ch). These reports are then of potential use to conservation planners, managers, and other decision makers to anticipate biodiversity losses in alpine and other systems across local, regional, and larger scales. Thus, an assessment is needed as to whether studies conducted at different scales yield comparable and consistent predictions.

A particular aspect that needs to be assessed is whether studies conducted at different spatial resolutions in a common study area and pool of species lead to different predictions of climate impacts. Such a discrepancy might, for instance, arise because different resolutions can reflect and represent topography, habitat, and climate variation differently. The relationship between certain environmental variables and species occurrence can differ when calculated at local or European scales (Guisan & Thuiller, 2005). For example, the mean temperature interpolated from local stations at a 20 m resolution (Dirnböck *et al*., 2003) contains more variability than expressed by the mean temperature within a 50 km × 50 km grid cell in which variation in elevation is poorly represented (Thuiller *et al*., 2005). Such differences in resolution and study area size might produce differences among SDMs in regard to the estimated minimal annual temperature tolerated by a species. This could happen when models are fitted with temperature data coming from relatively large cell sizes over a large area (e.g. Europe) that do not represent well substantial temperature variation with elevation. This, in turn, could result in underprediction of species distributions in areas at the cold end of the temperature gradient. In contrast, models fitted at a local scale and fine resolution could predict the persistence of suitable thermal habitat at high elevations within coarser resolution cells predicted overall to be unsuitable from models fitted at larger scale (e.g. from Thuiller *et al*., 2005).

Here, we examine this 'local high-elevation habitat persistence hypothesis.' We estimate, with fine-resolution (local-scale) models, the proportion of suitable habitat remaining for each species after climate change within cells predicted by coarse-resolution (European-scale) models to become overall unsuitable. We further investigate whether such predicted local habitat persistence results from (1) the particular topographic configuration in the local study areas; (2) the differences in estimated climatic niche of species when measured at both scales by determining whether the curves of probability of species presence along climatic gradients are truncated when fitted with local-scale data; (3) the choice of modeling techniques; and (4) difference in performance of models that are fitted under current conditions using data at the two extents and resolutions.

Methods

Study areas

We fitted SDMs for Western Europe (34°N–72°N, 11°W–32°E; 5.7×10^6 km^2) and for two local study areas in the

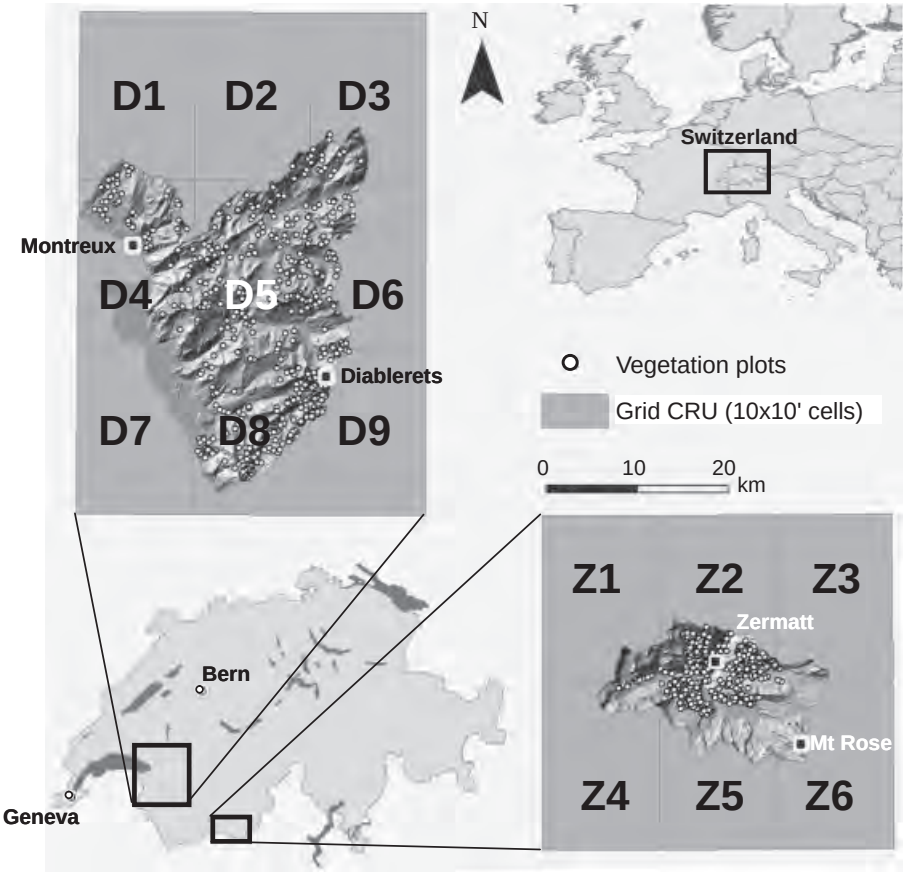


Fig. 1 Location of the Diablerets and Zermatt study areas. $10 \times 10'$ pixels used for projections at the European scale are indicated for each study area.

Western Alps: Diablerets and Zermatt (Fig. 1). The Diablerets study area (Fig. 1) covers nearly all mountain massifs of the Western Alps of the Canton de Vaud (Swiss state, 6°50′–7°10′E, 46°10′–46°30′N, $>700\,km^2$). The elevation ranges from 375 m in Montreux to 3210 m on the top of the Diablerets massif. The annual mean temperature and total precipitation vary, respectively, from 8 °C and 1200 mm at 600 m elevation to −5 °C and 2600 mm at 3000 m elevation (Bouët, 1985).

The Zermatt study area is located in the Central Alps of the Canton of Valais (Switzerland; 7°58′–7°91′E, 45°92′–46°06′N, 243 km^2) at the end of the Matter valley. Its elevation varies from 1480 m at the bottom of the valley near the village of Zermatt to 4634 m on the top of the Mt Rose massif. The climate conditions in the Matter valley have a continental character with low precipitation and high-radiation budgets. The annual mean temperature and total precipitation range, respectively, from 3.5 °C and 612 mm at 1638 m elevation to −6.5 °C and 680 mm at 3130 m elevation (MeteoSwiss). This area also holds glaciers that currently comprise 43% of the study area.

Species data

Data on the distributions of species in Europe were extracted from the AFE (Lahti & Lampinen, 1999), which uses a mapping grid of 50 km × 50 km cells (hereafter called 'AFE cells') based on the Universal Transverse Mercator (UTM) projection and the Military Grid Reference. Our sample included 2089 AFE cells. Species distributions data at the local scale were based on 550 64 m^2 vegetation plots in the Diablerets area and 1511 vegetation plots between 10 and 30 m^2 in the Zermatt area. Plots were restricted to open nonwoody vegetation (grassland, meadow, rock, and scree vegetation).

We developed models for 78 species that occur in more than 19 AFE cells and in at least one of the two local study areas. Of these species, 42 occur in the Diablerets and 51 in Zermatt (see Supporting information Appendix S1).

Climatic predictors

We used seven climatically derived variables expected to have a major direct ecophysiological impact on plant

species (see Prentice *et al*., 1992 for examples of variables; see Körner, 2003 for autoecology of alpine plants): growing degree days (GDD > 5 °C), mean annual temperature, minimum temperature of the coldest month, mean annual, winter and summer precipitation, and potential evapotranspiration. For the European scale, these climatic variables were obtained from the Climatic Research Unit (http://www.cru.uea.ac.uk) at a 10′ resolution. Mean values were averaged for the standard period 1961–1990. These $10 \times 10'$ maps were then aggregated to 50 km × 50 km resolution to match the AFE species data and allow fitting the models, which were then projected back on the $10 \times 10'$ maps. The Diablerets and the Zermatt study areas were captured by nine and six $10 \times 10'$ cells, respectively (Fig. 1).

We generated an identical set of environmental predictors for the Diablerets and Zermatt study areas at a local-scale resolution of 25 m (hereafter called '25 m × 25 m cells'). We first calculated linear lapse rates (i.e. rate of change along elevation) for long-term (1961–1990) monthly mean temperature and monthly rainfall taken from the national meteorological networks of Switzerland (MeteoSwiss). Next, we normalized the monthly values to sea level (0 m a.s.l.), using the regression lapse rates fitted along the elevation gradient, and interpolated the 0 m data to the whole surface of both study areas using inverse distance weighted interpolations (IDW). Finally, the spatially interpolated values (representing locally adjusted regression intercepts) were re-projected to actual elevations using a 25 m DEM (Digital Elevation Model) and the regression lapse rates (for details, see Zimmermann & Kienast, 1999). Additionally, the spatially distributed hydrological model PREVAH (Gurtz *et al*., 1999; Randin *et al*., 2006) was used to obtain a physically based predictor for potential evapotranspiration in both study areas, taking into account the effect of local topography.

Climate change scenarios

Spatial climate change projections were derived on the European scale for the 2080 time period ($10 \times 10'$ climatic grids; averages for 2070–2099) from the general circulation model (GCM) provided by the UK Hadley Center for Climate Prediction and Research (HadCM3; Carson, 1999), for which output from four different socioeconomic storylines (A1FI, A2, B1, B2) provided by the IPCC (Nakicenovic & Swart, 2000) were available. With an average increase of $+6.3 \pm 0.3$ °C in the two study areas for the period 2070–2099, the A1FI climate change scenario is the most extreme. B1 is mildest ($+3.2 \pm 0.2$ °C), and A2 ($+4.9 \pm 0.3$ °C) and B2 ($+3.5 \pm 0.2$ °C) are intermediate.

Climate change projections at the local scale were obtained for temperature and precipitation and the four IPCC scenarios by calculating monthly mean anomalies between the standard period 1961–1990 and the future time period 2070–2099 based on the $10 \times 10'$ climate grids. These anomalies were then downscaled to the 25 m resolution of local models using bilinear interpolation and added to the local-scale climatic predictors.

SDMs and their projections

We calibrated and projected SDMs (Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005) at both European and local scales using the latest version of the BIOMOD package (Thuiller, 2003), so as to follow as closely as possible the methods of Thuiller *et al*. (2005). For each species, generalized linear models (GLM), generalized additive models (GAM), and gradient boosting machine (GBM) were calibrated on a random sample of 70% of the initial observation data and evaluated on the remaining 30% dataset (but see Araujo *et al*., 2005b) using the area under the curve (AUC) of a receiver operating characteristic (ROC) plot (Fielding & Bell, 1997). We selected for each species and each scale the modeling technique that gives the best AUC value [best modeling technique approach – see Thuiller (2003)].

The most commonly used models for predicting species distribution so far are GLMs (e.g. Hill & Caswell, 1999; Bakkenes *et al*., 2002; Guisan *et al*., 2002) and GAMs are increasingly used (Yee & Mitchell, 1991; Frescino *et al*., 2001; Guisan *et al*., 2002; Thuiller *et al*., 2006a), whereas GBM (Friedman *et al*., 2000) has only recently been used (Leathwick *et al*., 2006; Broennimann *et al*., 2007; Pearman *et al*., 2008) and implemented in BIOMOD (Leathwick *et al*., 2006; Thuiller *et al*., 2006b). GBM was ranked as the best performing techniques in a recent large comparative analysis by Elith *et al*. (2006).

Models for 78 species were fitted at the European scale using the AFE distribution data at a resolution of 50 km × 50 km. These models were then used to predict current and future species' occurrence within the $10 \times 10'$ cells overlaying the two study areas, following the procedure proposed by Araujo *et al*. (2005a). Although we agree that this downscaling procedure can generate additional uncertainty, and thus should be used with caution, we used it for sake of comparability with previous studies at the European scale. At the local scale, models were fitted and species distributions were predicted for 42 species in the Diablerets and for 51 species in Zermatt using the 25 m × 25 m resolution climatic maps of the two study areas. Masks based on forests, lakes, urbanized areas, roads, and rivers were subsequently applied at the local scale to avoid spurious projections at locations that were not suitable

for reasons other than climate. Finally, species predictions were further restricted to landcover categories (grassland, meadow rock, and scree) on which the species was at least observed once.

For each species, we derived presence–absence predictions by using a threshold probability of presence that maximized the percentage of presences and absences correctly predicted in the training dataset (Pearce & Ferrier, 2000; Thuiller, 2003).

Calculating the persistence rate at local scale

A species would likely become extinct when predicted to lose 100% of its suitable habitat. However, because the link between habitat loss and extinction formally requires a population viability analysis (PVA) in addition to predictions of the spatial distribution of habitat (Botkin *et al*., 2007), we only discuss here our projections in terms of habitat loss. Four situations are possible when comparing predicted habitat suitability under current and future climatic conditions at the two scales (i.e. the four cells of a two-way contingency table). First, a $10 \times 10'$ cell might be suitable under the European-scale model and also contains suitable $25\,m \times 25\,m$ cells predicted by the local model. Second, a $10 \times 10'$ cell might be predicted suitable but have no suitable $25\,m \times 25\,m$ cell predicted within it. Third, a $10 \times 10'$ cell might be predicted unsuitable but one or more $25\,m \times 25\,m$ cells are predicted suitable within it. Finally, a $10 \times 10'$ cell might be predicted unsuitable and does not contain any suitable $25\,m \times 25\,m$ cells. Cases 2 and 3 represent mismatch in the modeling outcome at the two scales. Furthermore, loss of suitable habitat only occurs when a cell predicted suitable under current climate becomes unsuitable after climate change. Because our study areas only partially overlap with each $10 \times 10'$ cell, we can only quantify the third case here, allowing us to test the local scale 'refugia' hypothesis.

We examine how frequently local persistence is predicted to occur by calculating a coefficient across the s species predicted to be extinct within the $10 \times 10'$ cells. Let l_i be the total number of $10 \times 10'$ cells predicted to become unsuitable in the future for species i by the European model. Let then p_i be the number of times the local model predicts the persistence of at least one suitable $25\,m \times 25\,m$ cell for species i among the l_i $10 \times 10'$ cells. We calculated a local persistence coefficient P within each $10 \times 10'$ cell considered (cell-specific P) or for each of the s species separately considering all the $10 \times 10'$ cells they occurred initially (species-specific P) as follows:

$$P = \frac{\sum_{i=1}^{s} p_i}{\sum_{i=1}^{s} l_i}. \quad (1)$$

The local persistence coefficient reaches a maximum value of 100 when all species within the $10 \times 10'$ cells are predicted by local models to have sufficient habitat to persist.

Sensitivity analysis of the P *coefficient*

In order to discuss the potential errors generated by considering a minimum of 1 pixel in the P coefficient (e.g. the pixel may be a false-positive), P was calculated for a set of different minimum numbers of $25\,m \times 25\,m$ cells with suitable habitat remaining (1, 10, 50, 100, 200, 500, 1000, and 2000 $25\,m \times 25\,m$ cells).

Possible causes of habitat persistence at the local scale

We calculated several metrics to assess reasons for possible habitat persistence at the local scale. Predicted local persistence could result from differences in the elevation ranges represented in $10 \times 10'$ cells or from the divergence between the estimates of climatic conditions at the two scales. In the first step, we examined the temperature distribution of the $25\,m \times 25\,m$ cells within each $10 \times 10'$ cell. We then controlled whether a relationship existed between the persistence coefficient and (i) the percent overlay between the study area and the $10 \times 10'$ cells and (ii) the minimum, mean, maximum elevation, and range of elevation of $25\,m \times 25\,m$ cells within the $10 \times 10'$ cell and study area (i.e. the intersection). In the second step, we measured for each $10 \times 10'$ cell the agreement between the estimates of mean annual temperature and precipitation based on the low- and high-resolution climatic data available at the European and local scales, respectively. For this set of analysis, P was calculated separately for each $10 \times 10'$ cell by pooling the species (cell-specific P). For all following analyses, P was calculated separately for each species predicted extinct within the $10 \times 10'$ cells by pooling the $10 \times 10'$ cells (species-specific P).

Local persistence of habitat may also result from differences in species' elevation optimum, as high-elevation species mostly are expected to be negatively affected by climate change. To test this, we derived an index of elevation optimum for the 78 species as follows. A species was first assigned one of three possible values for its association (VA) with each elevation belt (EB from low to high: 1 = colline, 2 = montane, 3 = subalpine, 4 = alpine, 5 = nival), based on data from the Atlas Flora Alpina (Aeschimann *et al*., 2005). Values for VA are as follows: 0 if absent, 1 if not frequent, and 2 if commonly present within the corresponding elevation belt. The elevation optimum index was then calculated as

$$\text{Index} = \sum \frac{\text{EB} \times \text{VA}}{\sum \text{VA}}. \quad (2)$$

It resulted in four categories: 1 < montane ≤ 2; 2 < subalpine ≤ 3; 3 < alpine ≤ 4; 4 < nival ≤ 5. Because of the low number of nival species, these species were pooled with alpine species. We then examined the frequency with which species habitat persisted within each of these categories.

Likewise, local habitat persistence could depend on the position of species' optimum along climatic gradients. We therefore calculated the position of the environmental optimum of each species at both scales. Training data were concatenated, centered, and scaled for the seven climatic variables, so that the multidimensional space defined by the environmental variables was the same at the two scales. Finally, a principal component analysis (PCA) was applied to the difference between the centroids of the species' position at local and European scales (Pearman *et al*., 2008). The relationship between this difference and the persistence coefficient was tested with linear regressions under the four climate change scenarios.

Furthermore, truncated response curves may result in spurious future predictions of species' distributions (Thuiller *et al*., 2004). Therefore, we estimated the number of species with truncated response curves at local and European scales using Huisman–Olff–Fresco models (HOF models; Huisman *et al*., 1993; Oksanen & Minchin, 2002). HOF models include a hierarchical set of five models of increasing complexity: (I) flat curve with no response, (II) monotone increasing curve, (III) monotone increasing curve reaching a 'plateau,' (IV) symmetric unimodal curve, and (V) skewed unimodal response curve. Models I–III represent truncated responses, while models VI–V represent symmetric or skewed unimodal responses. We assessed the effect of a truncated response curve along the local and European temperature gradients on the persistence coefficient by testing whether species with truncated response curves along the local and/or European climatic gradients had a persistence rate different from that of species without truncated response curves (Wilcoxon's signed-rank tests).

Finally, we assessed the difference in model evaluation (AUC) for each species between the local and European scales. The relationship between the species' local persistence and the difference in AUC was estimated by a linear regression.

Results

Models performance

Models obtained on average a fair evaluation at the local scale (mean AUC = 0.84, SD = 0.08 for species in the Diablerets, Appendix S1-a; and mean AUC = 0.81, SD = 0.07 for species in Zermatt, Appendix S1-b) and a good evaluation at the European scale (mean AUC = 0.94, SD = 0.04 for species in the Diablerets, Appendix S1-a; and mean AUC = 0.96, SD = 0.04 for species in Zermatt, Appendix S1-b), making them useful for deriving future projections.

Table 1 Percentage of species predicted to become extinct within $10 \times 10'$ cells (C1–C15) at the European scales (ES) under the four climate change scenarios

$10 \times 10'$ cell	A1FI ES	A2 ES	B1 ES	B2 ES
C1	76.2	38.1	19.0	26.2
C2	61.9	26.2	9.5	23.8
C3	42.9	16.7	11.9	14.3
C4	83.3	47.6	31.0	28.6
C5	38.1	16.7	11.9	11.9
C6	38.1	16.7	11.9	14.3
C7	64.3	28.6	16.7	11.9
C8	38.1	11.9	11.9	14.3
C9	35.7	11.9	7.1	9.5
C10	52.3	6.8	13.6	4.5
C11	54.2	10.4	12.5	6.3
C12	40.5	2.4	11.9	2.4
C13	52.2	10.9	15.2	6.5
C14	45.5	2.3	13.6	2.3
C15	54.2	12.5	18.8	8.3
Mean	51.8	17.3	14.4	12.3
Standard deviation	14.5	12.8	5.6	8.3

Predicted species loss at the European scale

Predictions of species loss per $10 \times 10'$ cells at the European scale (Table 1) are within the same order as those predicted previously (Bakkenes *et al*., 2002; Thuiller *et al*., 2005). Depending on which $10 \times 10'$ cell is considered, maximum percent loss at the European scale was between 35.7% and 83.3% for the A1FI warming scenario and lowest rates between 2.3% and 28.6% for the B2 warming scenario (see also Supporting information Appendix S3).

Local persistence rate and its potential causes

Our results showed considerable variability in the temperature represented by 25 m × 25 m cells within each $10 \times 10'$ cell (Fig. 2). For all $10 \times 10'$ cells, the temperature at the European scale reflects the mean conditions at the local scale. However, low-temperature regions that represent potential refuges for plants are not captured at the coarse resolution of the European-scale data.

Local habitat persistence coefficients were high, ranging between 69% (A1FI) and 74% (B2) for the Diablerets area and 100% for the Zermatt area (Fig. 3; see also Supporting information Appendix S3). Difference in

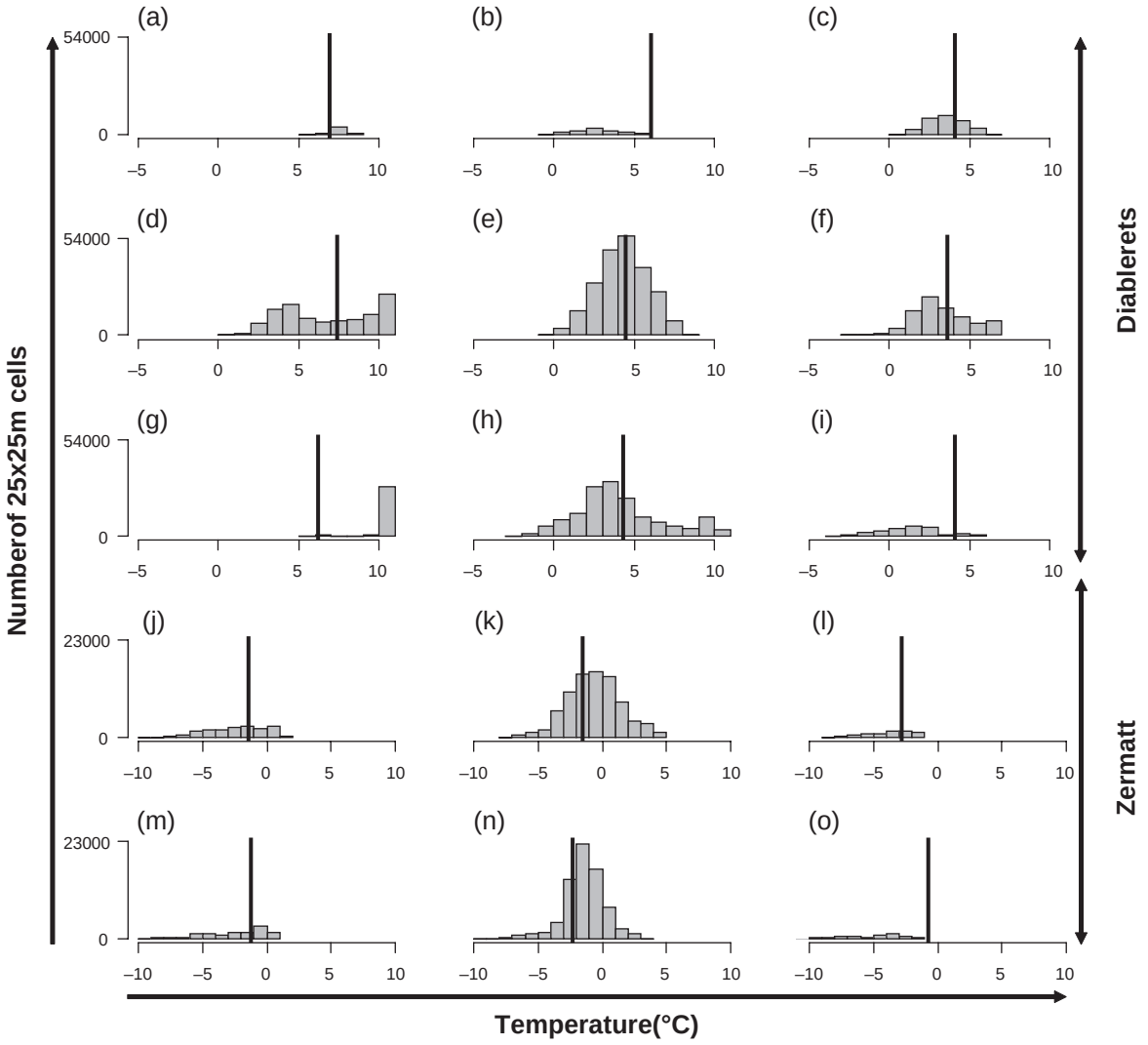


Fig. 2 Frequency distribution of the mean annual temperature (1 °C intervals) for the 25 m × 25 m cells within each 10 × 10′ cell of Diablerets (a to i: D1 to D9) and Zermatt (j to o: Z1 to Z6). The distribution of temperature reflects the local climatic variation within each 10 × 10′ cell. The vertical lines indicate the mean annual temperature of 10 × 10′ cells.

elevation range within 10 × 10′ cells was correlated with predicted local persistence. We found a strong and significant relationship between persistence rate and elevation range within 10 × 10′ cell under the four climate change scenarios (Fig. 4). The warmer the future climatic conditions, the higher the level of significance of the correlation. However, linear regressions showed no significant relationships between persistence rate and any other elevation attribute of 10 × 10′ cells (surface, minimum, mean, and maximum elevation; all *P*-values >0.05). In addition, the rate of habitat persistence per elevation belt was in general high for species with optima in the subalpine and alpine zones in both study areas (Table 2) but was the highest overall in Zermatt.

Finally, no significant relationship (*P*-value >0.05) was observed for any scenario between habitat persistence locally and any of the other factors tested: (i) difference in niche position, (ii) truncated response curves along temperature on one or both scales, (iii) model quality between the two scales, (iv) modeling techniques, and (v) model evaluation (AUC).

Sensitivity analysis of the P *coefficient*

When increasing the minimum number of remaining 25 m × 25 m cells from 1 to 2000, the *P* coefficient decreased by 19.9% and 19.1% under A1 for Diablerets and Zermatt, respectively (Fig. 5; see also Supporting information Appendix S4). The decrease under B2 was 3.1% for Diablerets, whereas the *P* coefficient remained stable in Zermatt.

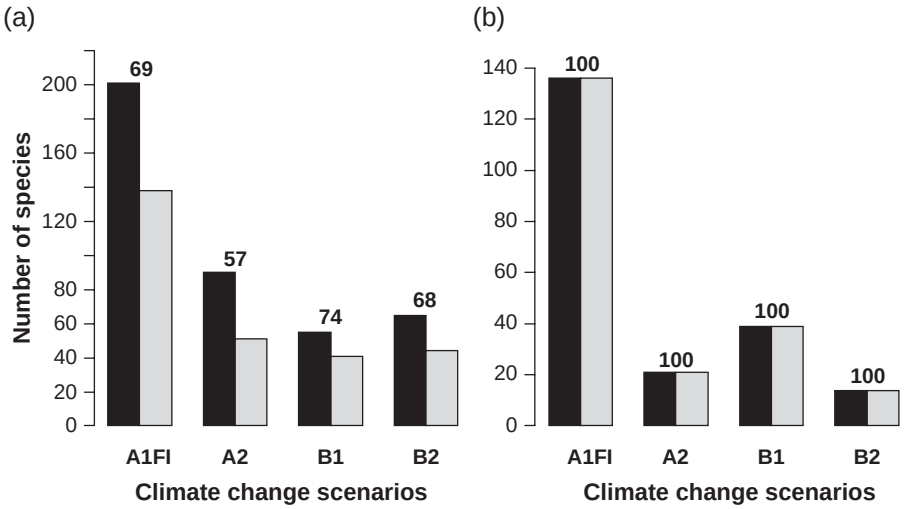


Fig. 3 Number of cases where a $10 \times 10'$ cell becomes unsuitable for a species (black) compared with the number of cases where suitable habitat persists for these species in a $10 \times 10'$ when modeled at the local (25 m × 25 m cells) scale (gray) for Diablerets (a) and Zermatt (b). The local persistence coefficient is indicated on the top of each bar. All $10 \times 10'$ cells in each study area are pooled.

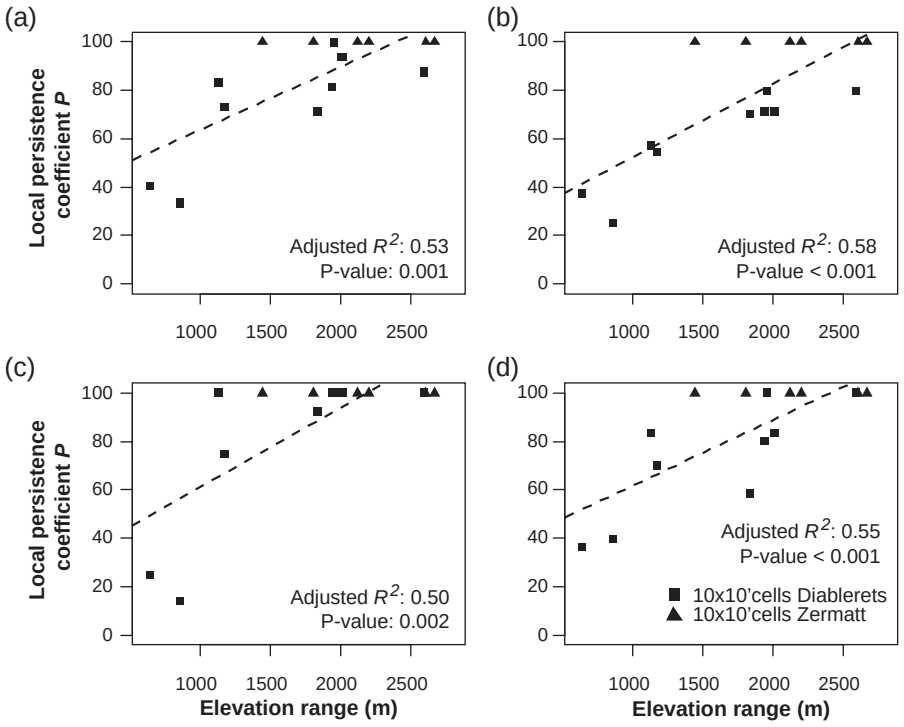


Fig. 4 Relationship between the local coefficient of persistence *P* and the elevation range of 25 m × 25 m cells within a $10 \times 10'$ cell under the four climate change scenarios (a: A1FI, b: A2, c: B1, d: B2). The *P* coefficient has been calculated for each $10 \times 10'$ cell individually (cell-specific *P*). The dashed lines represent regression trends.

Discussion

In this study, we assessed the 'local high-elevation persistence hypothesis' for 78 mountain species modeled at both European and local scales. We found that local models predicted persistence of some habitat of a number of species in $10 \times 10'$ cells that were predicted by models fitted at the European scale to contain no habitat. Persistence rates of habitat were especially high when considering the most severe warming scenario A1FI and were also greater in the Zermatt study area than in the Diablerets area owing to the presence of

higher elevations in the Zermatt area. Finally, the divergence between scales remained of ca. 50% when considering a minimum surface of ca. 1 km^2 for species to persist under the future climate conditions. Hence, our results give support to the persistence high-elevation habitat hypothesis, but too many limitations affecting the comparison between scales still prevent a formal testing of this hypothesis. Furthermore, it is impossible without data on future plant distributions to know if local projections are better than global ones.

Overall, we observed a strong relationship between the proportion of species with persisting habitat and the observed elevation range calculated from fine-grained data within $10 \times 10'$ cells. Greater habitat persistence also occurred in alpine species in Zermatt than in the Diablerets, suggesting that elevation range is the main driver for the predicted local-scale habitat persistence. None of the other relationships we explored showed a significant relationship with the proportion of species with persisting habitat.

Our results highlight the importance of assessing the potential impacts of climate change on species distributions at several scales, especially at local scale in mountain areas where the rugged topography requires fine mapping of environmental predictors. Possible high-elevation refuges for alpine and nival plants are likely better captured at a 25 m × 25 m resolution than at a $10 \times 10'$ resolution, as the latter corresponds to 16 km × 16 km cells in the Swiss Alps. This contrast between scales might help to explain the quaternary conundrum, i.e. why fewer species than expected went extinct during glacial periods when models predict so many extinctions with similar amplitude of climate change (Botkin *et al.*, 2007).

However, local predictions are also entailed with a major weakness compared with predictions at the extent of Europe: the full realized niche of species may be captured incompletely at local scale owing to the limited geographic – and therefore environmental – extent

Table 2 Persistence coefficient (*P*) per optimum of elevation

	Diablerets			Zermatt	
	M	S	A	S	A
A1FI	100	68	61	100	100
A2	68	73	17	100	100
B1	100	78	50	100	100
B2	77	81	23	100	100

The climate change scenarios are represented for each study area and $10 \times 10'$ cells in each study area are pooled (cell-specific *P*).
M, montane; S, subalpine; A, alpine.

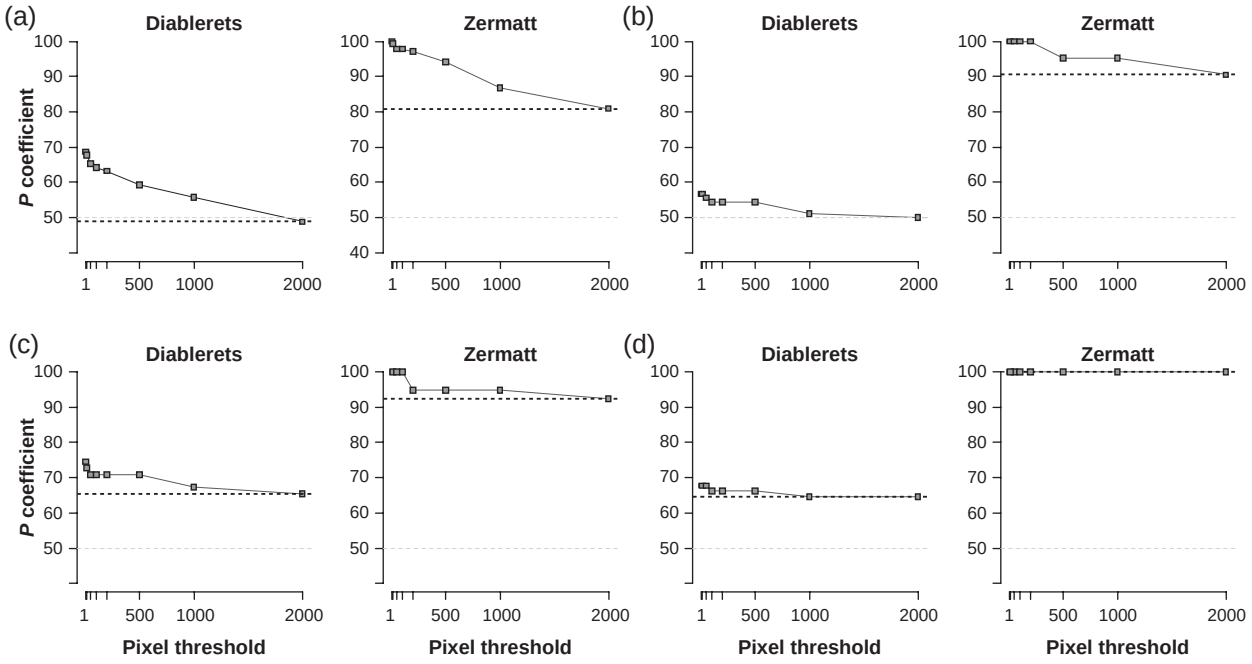


Fig. 5 Habitat persistence coefficient (*P*) calculated as the number of 25 m × 25 m cells with suitable habitat remaining for each species (1, 10, 50, 100, 200, 500, 1000, and 2000 25 m × 25 m cells) for each climate change scenario (a: A1FI, b: A2, c: B1 and d: B2) and for Diablerets and Zermatt (cell-specific *P* grouped by study area). The red dotted horizontal line represents the value of *P* when calculated with a minimum number of 25 m × 25 m cells necessary for a species to persist set to 2000, and the gray dotted horizontal line represents a *P* of 50 (50% of divergence between scales).

considered. In this case, truncated response curves may result for some species (especially low-elevation ones), and contribute to spurious predictions of future species distributions. If no distribution limit is established at both ends of the temperature gradient, species could potentially be predicted to either migrate indefinitely to higher elevations or remain present at the lowest elevation, even under climate change (Van Horn, 2002; Thuiller *et al.*, 2004). Indeed, although we did not observe any relationship between truncation of these curves and habitat persistence in our study, truncation can have effects on projections of species future distributions (e.g. Thuiller *et al.*, 2004) and thus deserves further investigation.

Some further limitations are also associated with local models. In our study, the set of local predictor variables may have not covered the full range of ecological requirements of the species, because it did not include process-based geomorphic predictors such as snow accumulation, rockfall, avalanche paths, or human-induced perturbations such as land use practices, which further contribute to limit species distributions (Dirnböck *et al.*, 2003; Dirnbock & Dullinger, 2004). However, we did not consider the retreat of glaciers in Zermatt, which in turn could contribute to increase available space for high-elevation species, and thus to further increase the divergence between scales. Absences may also not be comparable between the two scales. At local scale, absences may be due to a number of nonclimate-related factors, such as competitive exclusion, demographic processes, or environmental stochasticity, whereas climate is more likely to drive absences at the European scale.

Overall, the main reason for habitat persistence under the local scenario seems to be the importance of the length of elevation gradient, and probably the associated local topographic diversity, expressed within each large $10 \times 10'$ cell. Climatic differences along elevation gradients, as apparent at $25\,m \times 25\,m$ resolution (Fig. 2; see also Supporting information Appendix S2), allow plant species to find suitable climatic conditions at higher elevation under climate change. In contrast, models at a $10 \times 10'$ resolution reflect the mean climatic conditions within the cell, and thus provide imprecise values of the probability of occurrence of species along a thermal gradient. Global circulation models, from which some European-scale predictors were derived, likely do not express hygric continentality from the precipitation regime. This is suggested by precipitation being over- or underestimated in both study areas (results not shown, see Supporting information Appendix S2), thus contributing to the discrepancies between predictions at the two scales.

Continentality is an important climatic driver in mountain systems (Beniston, 2006). Inaccurate estimation of it by global or regional circulation models will hamper the prediction of species distributions in response to climate change. As highlighted by Nogués-Bravo *et al.* (2006), the coarse spatial resolution of GCMs/RCMs used at the European scale does not enable to capture the complex, topographically driven spatial patterning of temperature and other regional climate features at local scale. We thus propose a combination of GCM/RCM anomaly mapping downscaled and combined with finer scale present day climate maps, in order to better reflect local patterns of climate change within individual mountain ranges. Subalpine and alpine species in Zermatt were predicted to be less threatened by climate change – and thus show a higher persistence rate – than in the Diablerets. This result further highlights the importance of assessing the impact of climate change locally and independently for distinct mountain regions.

Local persistence may also result in local adaptations of species regarding their environment (e.g. ecotypes) and to local changes in biotic interactions that may result in changes of a species' realized niche. This may affect comparisons between European- and local-scale models and projections (Randin *et al.*, 2006).

Several new developments may contribute to the improvement of local models. First, our predictions were made assuming unlimited dispersal from present to future conditions found at higher elevations. Considering dynamic dispersal as climate changes progressively – and possibly nonlinearly over time – is an important constraint that needs to be added to static projections from SDMs (Thuiller *et al.*, 2008). So far, such examples only exist for few species (Carey, 1996; Dullinger *et al.*, 2004).

Second, the set of climatic predictors used in our study may be further improved to better express the true habitat requirements of species. Although our set of predictors was chosen to reflect as much as possible the known physiological requirements of species, deriving even more proximal predictors would allow a finer and ecologically more meaningful characterization of the species' realized niche (Guisan & Zimmermann, 2000; Austin, 2002). This should improve local projections of future species' distributions (Guisan & Thuiller, 2005).

Third, improved projections are likely to be obtained by combining the strength of models fitted at the two scales, by fitting the full realized climatic niche of a species from large-scale data picturing the whole species' range (e.g. AFE data at the European scale), and then refining the part of the niche where projections have to be made with finer scale data and local predictors. Such hierarchical approach of environmental drivers of species distributions (Pearson & Dawson, 2004) is worth further development and should be specifically tested for deriving improved projections of

future species distributions. Models could also be improved by considering additional techniques in an ensemble modeling approach (Araujo & New, 2007), like random forest (Prasad *et al*., 2006), support vector machine (Drake *et al*., 2006) or maxent (Phillips *et al*., 2006), which showed promising results in other modeling studies of species distribution.

Fourth and the last, the rough projections obtained from niche-based models could be used in conducting PVA for species in the study areas. Currently, this step cannot be made because we lack an adequate procedure to estimate extinction risks from the area predicted to remain suitable for a species under future climatic conditions (Botkin *et al*., 2007). In this study, we considered that a species could persist at the local scale if only one 25 m × 25 m cell persisted. Long-term persistence of species in only one such cell is ecologically questionable. However, even when considering a minimum viable surface of more than 1 km^2 for persistence (i.e. 2000 25 m × 25 m cells), the habitat persistence coefficient remained large (i.e. over 50%; see Supporting information Appendix S4). Adding a PVA requires, at least, information on (i) the degree of occupancy of suitable habitat by each species under current condition, (ii) the degree of contiguity of the remaining habitat, and (iii) the minimum absolute number of suitable cells a species requires to maintain positive population growth. This is a challenging, yet feasible, task (Morris & Doak, 2003).

Interestingly, a recent cross-scale comparison study found results opposite to ours (Trivedi *et al*., 2008). Projections of European-scale model (50 km × 50 km) predicted the persistence of 10 species in a mountain range of Scotland, while local models (50 m × 50 m) predicted the extinction of nine of them. The authors discuss that European models overestimated species' thermal tolerances, because the input coarse-resolution climate data were biased against the cold, high-altitude habitats of mountain plants. Further studies are thus required to assess, across a larger number of mountain ranges, whether local predictions over- or underpredict species extinctions compared with large-scale projections.

Conclusion

Our study yielded two main conclusions. First, local-scale models can predict persistent species habitat at high elevations within large cells that are predicted by coarse-resolution, European-scale models to become unsuitable (the 'local high-elevation refuge hypothesis'). Hence, European-wide projections might overestimate extinction risks for alpine species. Yet, for some species, local habitat persistence often owed to very few suitable 25 m × 25 m cells, suggesting a tenuous connection of habitat persistence with species persistence. Thus, for these species, predictions based on the European-scale data and resolutions remain plausible. These results require further testing. Secondly, models fitted at both scales examined here have strengths and drawbacks. In future studies, we suggest combining their strengths in a hierarchical approach that can estimate the full realized climatic niche of species while benefiting from finer environmental predictors locally.

Acknowledgements

This study was supported by the Centre de Conservation de la Nature (SFFN) of the Canton de Vaud, the Swiss National Science Foundation (SNF Grant No. 3100A0-110000), and the European Science Commission (FP6 projects ECOCHANGE and MACIS). We thank all the people who helped in the field, particularly Stéfanie Maire, Dario Martinoni, Séverine Widler, Chantal Peverelli, Valentine Hof, Grégoire Vuissoz, Lorenzo De Stefani, Roxane Milleret, Pierre Bermane Favrod-Coune, Daniel Cherix, Anny. and Willy Berra. C.R. and A.G. greatly benefited from the interactions with colleagues within the Swiss National Centre for Competence in Research (NCCR) 'Plant survival in natural and agricultural ecosystems' (http://www.unine.ch/nccr). We are particularly grateful to Dr Stefan Dullinger for his useful comments during the revision.

References

Aeschimann D, Heitz C (1996) *Index Synonymique de la Flore de Suisse et Territoires Limitrophes*. CRSF, Genève, Switzerland.

Aeschimann D, Lauber K, Moser DM *et al*. (2005) *Flora Alpina*. Belin, Paris, France.

Araujo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.

Araujo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.

Araujo MB, Thuiller W, Williams PH, Reginster I (2005a) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, **14**, 17–30.

Araujo MB, Whittaker RJ, Ladle RJ, Erhard M (2005b) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, **14**, 529–538.

Austin MP (2002) Case studies of the use of environmental gradients in vegetation and fauna modelling: theory and practice in Australia and New Zealand. In: *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds Scott JM, Heglund PJ, Samson F, Haufler J, Morrison M, Raphael M, Wall B), pp. 73–82. Island Press, Covelo, CA, USA.

Bahn M, Körner C (2003) Recent increase in summit flora caused by warming in the Alps. In: *Alpine Biodiversity in Europe* (eds Nagy L, Grabherr G, Körner C, Thompson DBA), pp. 437–441. Springer, Berlin, Germany.

Bakkenes M, Alkemade JRM, Ihle F, Leemans R, Latour JB (2002) Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology*, **8**, 390–407.

Beniston M (2006) Mountain weather and climate: a general overview and a focus on climatic change in the Alps. *Hydrobiologia*, **562**, 3–16.

Beniston M, Fox DG, Adhikary S *et al*. (1996) Impacts of climate change on mountain regions. In: *Intergovernmental Panel on Climate Change. Second Assessment Report* (ed. IPCC). pp. 191–213. Cambridge University Press, Cambridge, UK.

Botkin DB, Saxe H, Araujo MB *et al*. (2007) Forecasting the effects of global warming on biodiversity. *Bioscience*, **57**, 227–236.

Bouët M (1985) *Climat et météorologie de la Suisse romande*. Payot, Lausanne, Switzerland.

Braun-Blanquet J (1957) Ein Jahrhunder Florenwandel am Piz Linard (3414 m). *Bulletin du Jardin Botanique de l'Etat a Bruxelles*, volume jubilaire Walter Robyns, 221–232.

Broennimann O, Treier UA, Muller-Scharer H, Thuiller W, Peterson AT, Guisan A (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*, **10**, 701–709.

Carey PD (1996) DISPERSE: a cellular automaton for predicting the distribution of species in a changed climate. *Global Ecology and Biogeography Letters*, **5**, 217–226.

Carson DJ (1999) Climate modelling: achievements and prospects. *Quarterly Journal of the Royal Meteorological Society*, **125**, 1–27.

Davis AJ, Jenkinson LS, Lawton JH, Shorrocks B, Wood S (1998) Making mistakes when predicting shifts in species range in response to global warming. *Nature*, **391**, 783–786.

Diaz HF, Grosjean M, Graumlich L (2003) Climate variability and change in high elevation regions: past, present and future. *Climatic Change*, **59**, 1–4.

Dirnbock T, Dullinger S (2004) Habitat distribution models, spatial autocorrelation, functional traits and dispersal capacity of alpine plant species. *Journal of Vegetation Science*, **15**, 77–84.

Dirnböck 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.

Dormann CF (2007) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, **8**, 387–397.

Drake JM, Randin C, Guisan A (2006) Modelling ecological niches with support vector machines. *Journal of Applied Ecology*, **43**, 424–432.

Dullinger S, Dirnbock T, Grabherr G (2004) Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. *Journal of Ecology*, **92**, 241–252.

Elith J, Graham CH, Anderson RP *et al*. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecology*, **29**, 129–151.

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.

Frescino TS, Edwards TC, Moisen GG (2001) Modeling spatially explicit forest structural attributes using generalized additive models. *Journal of Vegetation Science*, **12**, 15–26.

Friedman JH, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, **28**, 337–374.

Gehrig-Fasel J (2007) *Treeline and climate change: analyzing and modeling patterns and shifts in the Swiss Alps*. PhD thesis, Université de Lausanne, Lausanne, 132 pp.

Grabherr G, Gottfried M, Pauli H (1994) Climate effects on mountain plants. *Nature*, **369**, 448.

Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89–100.

Guisan A, Holten JI, Spichiger R *et al*., (eds) (1995) *Potential Ecological Impacts of Climate Change in the Alps and Fennoscandian Mountains*. Conservatoire et Jardin Botaniques, Geneva, Switzerland.

Guisan A, Theurillat J-P (2000) Assessing alpine plant vulnerability to climate change: a modeling perspective. *Integrated Assessment*, **1**, 307–320.

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Gurtz J, Baltensweiler A, Lang H (1999) Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountain basins. *Hydrological Processes*, **13**, 2751–2768.

Hampe A (2004) Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, **13**, 469–476.

Hill MF, Caswell H (1999) Habitat fragmentation and extinction threshold on fractal landscapes. *Ecology Letters*, **2**, 121–127.

Hofer HR (1992) Veränderungen in der Vegetation von 14 Gipfeln der Berninagebietes zwischen 1905 und 1985. *Berichte des Geobotanischen Institutes der Eidgenössischen Technischen Hochschule Stiftung Rübel Zürich*, **58**, 39–54.

Huisman J, Olff H, Fresco LFM (1993) A hierarchical set of models for species response analysis. *Journal of Vegetation Science*, **4**, 37–46.

Hutchinson GE (1957) Population studies – animal ecology and demography – concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

Kearney M, Porter WP (2004) Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology*, **85**, 3119–3131.

Körner C (2003) *Alpine Plant Life*, 2nd edn. Springer, Berlin, Germany.

Lahti T, Lampinen R (1999) From dot maps to bitmaps – Atlas Flora Europaeae goes digital. *Acta Botanica Fennica*, **162**.

Leathwick JR, Elith J, Francis MP, Hastie T, Taylor P (2006) Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology – Progress Series*, **321**, 267–281.

Morris WF, Doak DF (2003) *Quantitative Conservation Biology: Theory and Practice of Population Viability Analysis*. Sinauer Associates, Sunderland, MA, USA.

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, UK.

Nogués-Bravo D, Araújo MB, Martinez-Rica JP, Errea MP (2006) Exposure of global mountain systems to climate warming during the 21st century. *Global Environmental Change*, **17**, 420–428.

Oksanen J, Minchin PR (2002) Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling*, **157**, 119–129.

Pauli H, Gottfried M, Grabbherr G (1996) Effects of climate change on mountain ecosystems. Upward shifting of alpine plants. *World Resource Review*, **8**, 382–390.

Pauli H, Gottfried M, Reiter K, Klettner C, Grabherr G (2007) Signals of range expansions and contractions of vascular plants in the high Alps: observations (1994–2004) at the GLROIA master site Schrankogel, Tyrol, Austria. *Global Change Biology*, **13**, 147–156.

Pearce J, Ferrier S (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127–147.

Pearman P, Van der Knaap P, Randin CR *et al*. (2008) Prediction of plant species distributions across six millennia. *Ecology Letters*, **11**, 357–369.

Pearson D, Dawson TP (2004) Modelling species distribution in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285–298.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, **9**, 181–199.

Prentice C, Cramer W, Harrison SP, Leemans R, Monserud RA, Solomon AM (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, **19**, 117–134.

Randin CF, Dirnbock T, Dullinger S, Zimmermann NE, Zappa M, Guisan A (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.

Theurillat J-P, Felber F, Geissler P *et al*. (1998) Sensitivity of plant and soil ecosystems of the Alps to climate change. In: *Views from the Alps. Regional Perspectives on Climate Change* (eds Cebon P, Dahinden U, Davies HC, Imboden D, Jäger CC), pp. 225–308. MIT Press, London, UK.

Thomas CD, Cameron A, Green RE *et al*. (2004) Extinction risk from climate change. *Nature*, **427**, 145–147.

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, Albert C, Araújo MB *et al*. (2008) Predicting global change impacts on plant species distributions: where to go from here? *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137–152.

Thuiller W, Broennimann O, Hughes G, Alkemade JRM, Midgley GF, Corsi F (2006a) Vulnerability of African mammals to anthropogenic climate change under conservative land transformation assumptions. *Global Change Biology*, **12**, 424–440.

Thuiller W, Brotons L, Araujo MB, Lovorel (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165–172.

Thuiller W, Lavorel S, Araujo MB, Sykes MT, Prentice IC (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8245–8250.

Thuiller W, Midgley GF, Rouget M, Cowling RM (2006b) Predicting patterns of plant species richness in megadiverse South Africa. *Ecography*, **29**, 733–744.

Trivedi MR, Berry PM, Morecroft MD, Dawson TP (2008) Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. *Global Change Biology*, **14**, 1089–1103.

Van Horn B (2002) Approaches to habitat modelling: the tensions between pattern and process and between specificity and generality. In: *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds Scott JM, Heglund PJ, Morrison ML *et al*.), pp. 63–72. Island Press, Covelo, CA, USA.

Vittoz P, Jutzeler S, Guisan A (2006) Flore alpine et réchauffement climatique: observation de trois sommets valaisans à travers le 20ème siècle. *Bulletin de la Murithienne*, **123/2005**, 49–59.

Walther G-R (2003) Plants in a warmer world. *Perspectives in Plant Ecology, Evolution and Systematics*, **6**, 169–185.

Walther G-R, Beißner S, Burga CA (2005) Trends in the upward shift of alpine plants. *Journal of Vegetation Science*, **16**, 541–548.

Yee TW, Mitchell ND (1991) Generalized additive models in plant ecology. *Journal of Vegetation Science*, **2**, 587–602.

Zimmermann NE, Kienast F (1999) Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science*, **10**, 469–482.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Species found in either (a) the Diablerets or (b) Zermatt dataset and are in the Atlas Florae Europaeae (AFE). The best modeling technique used for spatial projection and its area under the curve (AUC) value are indicated for each species and each scale. The elevation index (EI) is the last column. See separated Excel files for (a) the Diablerets and (b) Zermatt.

Appendix S2. Relationships between (a) projections of the mean annual temperature at the local scale (x-axis) and within the $10 \times 10'$ cells at the European scale (y-axis) and (b) projections of the mean annual precipitation at the local scale (x-axis) and within the $10 \times 10'$ cells at the European scale (y-axis). The mean temperature and mean precipitation at the local scale are calculated by averaging the values of all $25\,m \times 25\,m$ cells of one of the study areas within a $10 \times 10'$ cell. The diagonal represents a perfect agreement between scales and the dashed lines are regressions lines.

Appendix S3. Number of $25\,m \times 25\,m$ (LS) and $10 \times 10'$ (ES) cells remaining for each species under the four climate change scenarios for (a) the Diablerets (D1 to D9) and (b) Zermatt (Z1 to Z6). See separated Excel files for (a) the Diablerets and (b) Zermatt.

Appendix S4. Habitat persistence coefficient (P) calculated as the number of $25\,m \times 25\,m$ cells with suitable habitat remaining for each species (1, 10, 50, 100, 200, 500, 1000, and 2000 $25\,m \times 25\,m$ cells) for each climate change scenario.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix S1. Species found in either (a) the Diablerets or (b) Zermatt dataset and are in the Atlas Florae Europaeae (AFE). The best modeling technique used for spatial projection and its area under the curve (AUC) value are indicated for each species and each scale. The elevation index (EI) is the last column. Table S1a) Diablerets, Table S1b) Zermatt.

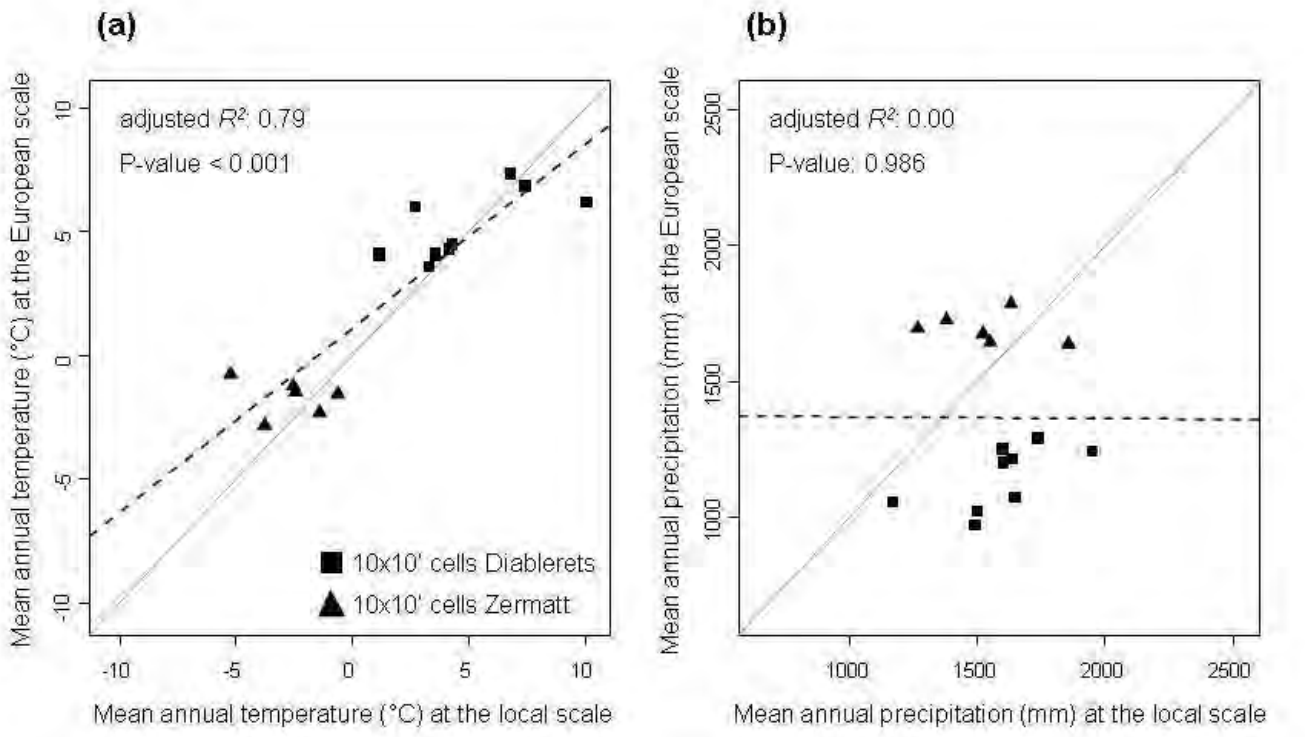
Table S1a)

Species	Best model LS	AUC LS	Best model ES	AUC ES	EI
Aconitum napellus aggr.	GLM	0.74	GBM	0.94	3.0
Arabis alpina s.str.	GLM	0.90	GAM	0.91	3.1
Arabis caerulea	GAM	0.97	GAM	0.98	4.0
Arabis hirsuta	GBM	0.82	GBM	0.82	1.8
Asplenium viride	GBM	0.85	GBM	0.86	2.7
Biscutella laevigata	GBM	0.68	GAM	0.94	2.8
Botrychium lunaria	GAM	0.77	GBM	0.91	2.8
Cardamine pratensis	GAM	0.92	GBM	0.93	1.8
Cerastium arvense s.l.	GAM	0.59	GBM	0.91	1.5
Cerastium fontanum subsp. vulgare	GLM	0.87	GLM	0.95	2.0
Cerastium latifolium	GLM	0.92	GAM	0.98	3.6
Cystopteris fragilis	GBM	0.87	GBM	0.82	2.7
Draba aizoides	GLM	0.85	GAM	0.94	3.2
Gypsophila repens	GLM	0.76	GAM	0.98	2.8
Juniperus communis subsp. nana	GLM	0.94	GBM	0.91	3.5
Picea abies	GAM	0.85	GBM	0.97	2.5
Polygonum bistorta	GAM	0.90	GBM	0.94	2.2
Polygonum viviparum	GLM	0.81	GLM	0.98	3.2
Polystichum lonchitis	GAM	0.84	GLM	0.90	2.5
Pritzelago alpina s.str.	GAM	0.89	GAM	0.98	3.6
Pulsatilla alpina s.str.	GLM	0.83	GAM	0.97	3.2
Ranunculus aconitifolius	GAM	0.78	GAM	0.98	2.5
Ranunculus acris s.l.	GBM	0.88	GAM	0.96	2.3
Ranunculus alpestris	GLM	0.85	GAM	0.98	3.2
Ranunculus bulbosus	GAM	0.94	GBM	0.92	2.2
Ranunculus montanus aggr.	GBM	0.83	GAM	0.95	2.7
Ranunculus repens	GBM	0.84	GBM	0.95	1.8
Rumex acetosa	GLM	0.94	GBM	0.96	1.8
Rumex alpestris	GLM	0.88	GBM	0.94	3.0
Rumex alpinus	GLM	0.84	GBM	0.93	3.0
Rumex crispus	GBM	0.90	GBM	0.93	1.8
Sagina saginoides	GBM	0.90	GLM	0.92	3.2
Salix reticulata	GAM	0.81	GAM	0.97	3.6
Salix retusa	GLM	0.84	GAM	0.99	3.6
Selaginella selaginoides	GBM	0.83	GLM	0.96	3.2
Silene acaulis	GAM	0.89	GLM	0.97	4.0
Silene dioica	GBM	0.78	GBM	0.93	2.2
Silene vulgaris s.l.	GBM	0.75	GBM	0.79	2.3
Stellaria graminea	GBM	0.86	GBM	0.93	1.8
Thesium alpinum	GAM	0.80	GAM	0.93	2.7
Trollius europaeus	GLM	0.72	GBM	0.92	2.5
Urtica dioica	GLM	0.68	GBM	0.91	2.0

Table S1b)

Species	Best model LS	AUC LS	Best model ES	AUC ES	EI
Arabis alpina s.str.	GBM	0.82	GAM	0.91	3.1
Arabis caerulea	GLM	0.83	GAM	0.98	4.0
Arabis ciliata	GAM	0.75	GAM	0.98	2.6
Arenaria ciliata	GAM	0.77	GLM	0.98	4.0
Biscutella laevigata	GAM	0.86	GAM	0.94	2.8
Botrychium lunaria	GAM	0.62	GBM	0.91	2.8
Cardamine resedifolia	GBM	0.80	GAM	0.96	3.7
Cerastium arvense subsp. strictum	GLM	0.73	GBM	0.91	2.7
Cerastium cerastoides	GBM	0.82	GAM	0.97	3.5
Cerastium latifolium	GLM	0.92	GAM	0.98	3.6
Cerastium pedunculatum	GAM	0.93	GAM	1.00	4.0
Cerastium uniflorum	GBM	0.91	GLM	0.99	4.2
Dianthus carthusianorum subsp. vaginatus	GBM	0.92	GBM	0.96	1.8
Dianthus sylvestris	GAM	0.94	GLM	0.96	2.3
Draba aizoides	GAM	0.77	GAM	0.94	3.2
Draba dubia	GAM	0.91	GLM	0.99	3.6
Draba fladnizensis	GLM	0.78	GBM	0.98	4.0
Draba hoppeana	GAM	0.92	GLM	1.00	4.0
Draba siliquosa	GAM	0.83	GLM	0.99	4.0
Equisetum variegatum	GBM	0.91	GBM	0.89	2.7
Erysimum rhaeticum	GAM	0.87	GLM	0.98	2.2
Gypsophila repens	GBM	0.83	GAM	0.98	2.8
Herniaria alpina	GBM	0.80	GLM	0.99	3.6
Juniperus communis subsp. nana	GAM	0.76	GBM	0.91	3.5
Juniperus sabina	GAM	0.89	GAM	0.96	2.3
Minuartia recurva	GLM	0.77	GAM	0.92	3.7
Minuartia sedoides	GLM	0.83	GAM	0.97	4.0
Minuartia verna	GBM	0.68	GBM	0.89	3.5
Oxyria digyna	GBM	0.76	GLM	0.96	4.0
Pinus cembra	GBM	0.84	GLM	0.99	3.0
Polygonum viviparum	GLM	0.75	GLM	0.98	3.2
Pritzelago alpina subsp. brevicaulis	GLM	0.84	GAM	0.98	3.8
Pulsatilla alpina subsp. apiifolia	GBM	0.79	GAM	0.97	3.0
Pulsatilla halleri	GAM	0.90	GAM	0.95	3.2
Pulsatilla vernalis	GLM	0.77	GBM	0.94	2.8
Ranunculus glacialis	GBM	0.87	GAM	0.96	4.2
Ranunculus montanus aggr.	GLM	0.77	GAM	0.95	2.7
Rumex acetosella s.str.	GAM	0.84	GBM	0.94	2.3
Sagina saginoides	GLM	0.77	GLM	0.92	3.2
Salix breviserrata	GLM	0.82	GLM	0.99	3.3
Salix foetida	GBM	0.76	GLM	1.00	3.3
Salix herbacea	GLM	0.72	GBM	0.97	4.0
Salix reticulata	GBM	0.80	GAM	0.97	3.6
Salix retusa	GBM	0.74	GAM	0.99	3.6
Salix serpillifolia	GBM	0.69	GLM	0.99	4.0
Selaginella selaginoides	GAM	0.79	GLM	0.96	3.2
Silene nutans s.str.	GLM	0.88	GBM	0.93	1.8
Silene rupestris	GBM	0.74	GAM	0.90	2.5
Silene vulgaris s.str.	GAM	0.92	GBM	0.79	2.3
Thalictrum foetidum	GAM	0.82	GLM	0.93	2.3
Thesium alpinum	GLM	0.74	GAM	0.93	2.7

Appendix S2. Relationships between (a) projections of the mean annual temperature at the local scale (*x*-axis) and within the 10 × 10′ cells at the European scale (*y*-axis) and (b) projections of the mean annual precipitation at the local scale (*x*-axis) and within the 10 × 10′ cells at the European scale (*y*-axis). The mean temperature and mean precipitation at the local scale are calculated by averaging the values of all 25 m × 25 m cells of one of the study areas within a 10 × 10′ cell. The diagonal represents a perfect agreement between scales and the dashed lines are regressions lines.



Appendix S3. Number of 25 m × 25 m (LS) and 10 × 10′ (ES) cells remaining for each species under the four climate change scenarios for (a) the Diablerets (D1 to D9) and (b) Zermatt (Z1 to Z6). See separated Excel files for (a) the Diablerets and (b) Zermatt.

http://www3.interscience.wiley.com/journal/121479028/suppinfo

Appendix S4. Habitat persistence coefficient (*P*) calculated as the number of 25 m × 25 m cells with suitable habitat remaining for each species (1, 10, 50, 100, 200, 500, 1000, and 2000 25 m × 25 m cells) for each climate change scenario.

http://www3.interscience.wiley.com/journal/121479028/suppinfo

A 4

Prediction of plant species distributions across six millennia.

Pearman P.B, Randin C.F., Broennimann O., Vittoz P., van der Knaap W.O., Engler R., Le Lay G., Zimmermann N.E. and Guisan A, 2008. *Ecology Letters*, **11**, 357-369.

Ecology Letters, (2008) **11**: 357–369 doi: 10.1111/j.1461-0248.2007.01150.x

LETTER

Prediction of plant species distributions across six millennia

Peter B. Pearman,[1]* Christophe F. Randin,[1] Olivier Broennimann,[1] Pascal Vittoz,[1] Willem O. van der Knaap,[2] Robin Engler,[1] Gwenaelle Le Lay,[1] Niklaus E. Zimmermann[3] and Antoine Guisan[1]

[1]Department of Ecology and Evolution, University of Lausanne-Biophore, CH-1015 Lausanne, Switzerland
[2]Institut für Pflanzenwissenschaft, Altenbergrain 21H, CH-3013 Bern, Switzerland
[3]Land Use Dynamics, Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
**Correspondence: E-mail: pearman@wsl.ch*

Abstract

The usefulness of species distribution models (SDMs) in predicting impacts of climate change on biodiversity is difficult to assess because changes in species ranges may take decades or centuries to occur. One alternative way to evaluate the predictive ability of SDMs across time is to compare their predictions with data on past species distributions. We use data on plant distributions, fossil pollen and current and mid-Holocene climate to test the ability of SDMs to predict past climate-change impacts. We find that species showing little change in the estimated position of their realized niche, with resulting good model performance, tend to be dominant competitors for light. Different mechanisms appear to be responsible for among-species differences in model performance. Confidence in predictions of the impacts of climate change could be improved by selecting species with characteristics that suggest little change is expected in the relationships between species occurrence and climate patterns.

Keywords

Climate change, global circulation model, hindcasting, Holocene, niche conservatism, PMIP, pollen, range filling, species distribution model.

Ecology Letters (2008) 11: 357–369

INTRODUCTION

The earth is currently experiencing rapid, anthropogenic climate change (Houghton *et al.* 2001) that is expected to impact species diversity, distribution and persistence (Thomas *et al.* 2004; Thuiller *et al.* 2005; Botkin *et al.* 2007). For example, the ranges of insects and birds are already expanding northward as more northerly areas become increasingly suitable (Walther *et al.* 2002; Parmesan & Yohe 2003). Similarly, plants in mountainous regions are responding by shifting elevation ranges upwards (Grabherr *et al.* 1994; Gehrig-Fasel *et al.* 2007). Potential effects of climate change on the distributions of species are often evaluated using niche-based species distribution models (SDMs), in which current climate and species distribution data are used to model the realized climatic niche (Hutchinson 1957). Niche models are then projected in geographic space using estimates of future climate patterns (Guisan & Zimmermann 2000; Guisan & Thuiller 2005). Models suggest that species with limited dispersal abilities and/or high-elevation habitats will become threatened with extinction as suitable habitat becomes reduced and new areas remain unreachable due to natural and anthropogenic barriers to dispersal (Hannah *et al.* 2002; Broennimann *et al.* 2006). Nonetheless, predicted changes in species distribution are difficult to evaluate with empirical data because the predicted changes may take decades to centuries to occur (Lang 1994; Araújo *et al.* 2005; Bradshaw & Lindbladh 2005; Araújo & Rahbek 2006).

Predictions from SDMs can be affected by factors such as data resolution, sampling extent and choice of modelling algorithm (Elith *et al.* 2006; Randin *et al.* 2006; Guisan *et al.* 2007a; Thuiller *et al.* in press). While such effects may be mitigated by careful design regarding these aspects, shifts in species niche that are caused by dynamic ecological or evolutionary processes could bias or invalidate predictions of the biotic effects of climate change obtained from SDMs (Pearman *et al.* 2008). Prediction using SDMs of species' future distributions assumes that species niches do not change over the relevant time scale. However, some plant species seem to have undergone rapid niche shifts (Broennimann *et al.* 2007) while other plants appear to experience long periods of niche stability (Huntley *et al.* 1989). The unassessed potential for niche shift casts doubt upon the

validity of SDM-based predictions of climate-change impacts (Pearman *et al.* 2008). While future niche changes are unlikely predictable for particular species, increased reliability of SDM-based predictions may depend on understanding how potentials for niche change vary among species. The consistent association of species characteristics with unchanging distribution–climate relationships would assist ecologists in determining the quality of predictions and increase confidence in projected climate-change impacts obtained from SDMs.

The modelling and prediction of climate-change impacts on species distributions additionally assumes that species have migrated to fill the distribution that is accessible given the environmental requirements of the species and the outcome of competitive interactions (Pearson & Dawson 2003). If this is not the case, it may be difficult to develop SDMs that accurately represent species' niches. A difference in time between changes in the geographic distribution of climatic conditions that are suitable for a species and the colonization of newly suitable areas by the species (or transient persistence of slowly declining populations in areas no longer suitable) could be indistinguishable from a shift in a species niche. Changes in the observed relationship between climate and species distributions could, thus, be generated by temporal changes in proportion occupancy of a species potential geographic range (Davis & Shaw 2001; Svenning & Skov 2004). For example, European beech (*Fagus sylvatica* L.) might currently occupy most of its potential range and be at distributional equilibrium (Huntley *et al.* 1989) while *Abies alba* L. could occupy only 37% of its potential range (Svenning & Skov 2004). Incomplete range filling currently might result from dispersal limitation of range expansion from Pleistocene refugia (Svenning & Skov 2004). If so, then partial range filling may have been even more pronounced in some species during the mid-Holocene. It follows that if such limitations have already lasted thousands of years, dispersal limitation would likely be substantial in response to rapid climate warming. This would clearly impede accurate projections of future plant distributions unless dynamic dispersal is implicitly taken into account in the modelling (Thuiller *et al.* in press). Nonetheless, the effects that niche shifts and partial range filling may have on predictions from niche-based SDMs have never been investigated using appropriate independent data and rigorous statistical methods.

In this paper, we test the predictive ability of SDMs using forecasting and hindcasting of species distributions as functions of past and current climates (Araújo & Rahbek 2006). Forecasting is the process of fitting statistical models using data on present climate and distributions, and then projecting species potential distributions into the future using estimations of future climate. Similarly, one might calibrate models using data on past climate and species distributions and then evaluate model forecasts for current species distributions by comparing the predictions to known current distributions. In contrast, hindcasting is the process by which one calibrates models using data on current climatic conditions and species distributions, and then projects the modelled relationships into the past using independent estimates of prior climates. So far, only a few studies have used hindcasting to estimate previous species distributions, for example, in testing for niche changes between last glacial maximum and the present (Martinez-Meyer & Peterson 2006). Here, we explore hindcasting as a method to assess quantitatively the accuracy of predictions of climate-change impacts on biodiversity and species distributions, and of how predictions of species distributions vary when the assumptions behind predictive application of SDMs are violated.

To assess model predictive performance during hindcasting and forecasting, one needs sufficient, independent data on current and historical species distributions, and on present and past climate. We use atlas data on current plant distributions, pollen core data from two European databases and climate estimates from a global circulation model (GCM) to conduct an independent assessment of SDM performance upon hindcasting the distributions of tree species at the mid-Holocene, 6 ky BP. Similarly, we forecast current distributions of these species by using pollen data on species' mid-Holocene distributions to calibrate (i.e. fit) the models. We use multivariate techniques to estimate change in the niche position (i.e. change in 'marginality' of species in multivariate climate space, *sensu* Dolédec *et al.* 2000) of each species between the mid-Holocene and the present. We then evaluate the relationship between these estimated niche shifts and model predictive performance. Models for species would not likely perform identically as a consequence of potential interspecific variation in range filling, niche stability and data quality. We quantify the uncertainty surrounding among-species differences in model performance by providing bootstrapped 95% confidence intervals. Finally, we interpret variation in model performance among species in terms of species ecological characteristics, interspecific competition and range expansion.

DATA AND METHODS

Species distributions

Current distributions of plant species in Europe were taken from the digital Atlas Florae Europaeae (AFE) database (Jalas & Suominen 1972–1999). We eliminated off-shore grid cells, leaving a total of 1973 cells with which we determined species presence/absence. To determine species distributions at 6 ky BP, we examined pollen composition in the pooled sample of cores in the Alpine Palynological

Database and the European Pollen Database (http://www.europeanpollendatabase.net). Pollen data included the interval 6 ± 0.5 ky BP, using calibrated C^{14} dates. Pollen from wetland and aquatic plants were removed from the data so that the data reflected each species' pollen as a proportion of pollen from terrestrial species. We then averaged pollen percentages for each species across the 10 100-year periods.

Direct species-level determinations are generally not possible for pollen morphotypes because pollen differs little among species within a genus. To allow SDMs of species, we selected only those types of pollen that represented a single species in northern and central Europe and thus could not be confused with pollen arising from other members of a genus. To achieve this, we restricted chosen pollen sites to be north of a line drawn from the Adriatic Sea north-eastward, passing through northern Romania and continuing eastward *c.* 200 km north of the Black Sea. This left an area of continental Europe, Scandinavia and the United Kingdom/Ireland, and removed cores where species common in central and northern Europe might overlap in range with other congeners. We used data from all remaining 312 cores to maximize opportunity for model calibration and testing (Table 1).

Once the study species were identified, we established two threshold pollen percentages per species, based on expert knowledge, for determining species presence/absence. A lower threshold per cent for species presence was established to signify a level below which we considered the species unlikely to be present. Likely this would capture relatively small and recently established populations, but species presence determinations might be influenced by long-distance pollen transport (see Discussion in Latalowa & van der Knaap 2006). An upper threshold per cent, if exceeded, led us to consider a species as definitely present, minimizing the influence of pollen transport. For each species, pollen percentages were evaluated using the average percentages over the 10 100-year period centred on 6 ky BP. Here, we present results using the lower threshold, based on the observation that plant macrofossils often indicate a date for species arrival that is earlier than pollen evidence (Kullman 2001; Magri *et al.* 2006). Because of this phenomenon, the use of our high pollen thresholds for model calibration and evaluation may be unnecessarily conservative and in our data sets some species showed an insufficient number of presences for reliable modelling. We present analyses using the upper threshold in supplementary online materials.

Current climate

We used interpolated climate data at 1-km resolution from the WorldClim data set (Hijmans *et al.* 2005), downloaded 31 March 2006. We chose to use the variables 'annual mean temperature', 'mean temperature of the coldest month', 'total annual precipitation', 'precipitation December–March' and 'precipitation June–August' because of the close relationship of these or very similar variables to plant physiological limitations (Bartlein *et al.* 1986; Prentice *et al.* 1992). In a geographic information system, we sampled each climate variable map at locations corresponding to the centre of each of the 1973 AFE cells.

Mid-Holocene climate estimate

The use of climate estimates reconstructed from pollen composition in cores (Davis *et al.* 2003) for hindcasting mid-Holocene plant distributions would have generated circularities in the results. This is because the same pollen data were used to determine species distributions and resulting

Table 1 Current species distributions (presence/absence) based on Atlas Floreae Europaeae (AFE) and mid-Holocene presence/absence based on pollen percentages from the combined holding of the European Pollen Database and Alpine Palynological Database

	Current			Mid-Holocene	
	AFE	Pollen thresholds*		Presences/absences	
Species	Presences/absences	Low	High	Low threshold	High threshold
Abies alba	410/1573	0.01	0.02	71/241	62/250
Carpinus betulus	1000/983	0.005	0.02	14/289	8/304
Corylus avellana	1423/560	0.01	0.03	198/114	145/167
Fagus sylvatica	990/993	0.01	0.02	47/265	31/281
Juniperus communis	1486/497	0.01	0.02	21/291	10/302
Larix decidua	135/1848	0.005	0.015	19/293	12/300
Picea abies	783/1200	0.01	0.02	119/193	93/219

*Employed threshold values necessary for species presence when considering mean pollen percentage over 10 consecutive 100-year time periods in which pollen of the genus was detected.

biome maps upon which reconstructed climate estimates were ultimately based. To avoid this circularity we estimated mid-Holocene climate using two data sets. To obtain an estimate of the European climate anomaly between the current period and the mid-Holocene, we obtained projected climate data from the UBRIS-HadCM3 AO GCM (Gordon *et al.* 2000), generated as a part of the PMIP2 climate modelling comparison project (Gladstone *et al.* 2005). Comparison of the results of this GCM with climate reconstructions and other models show that the direction of climate change is in general correctly estimated in the PMIP2 models, although the degree of cooling in southern Europe is generally underestimated (Brewer *et al.* 2007). The UBRIS-HadCM3 model results were available for both the present (*c.* 1950) and for a 100-year period centred around 6 ky BP. We obtained average values for each variable for each month over the 100-year period. We then generated an anomaly map for each variable by subtracting values for the present, as modelled by the GCM, from the GCM estimated mid-Holocene values. These anomaly maps were then interpolated to 1-km resolution using inverse-distance weighting among the four nearest cells in ARCGIS 9.1 (ESRI 2005). Values in these interpolated maps were then added to the corresponding WorldClim values to generate estimated climate maps at 1-km resolution for Europe at the mid-Holocene. From these maps, we extracted climate data at the 312 pollen core sites from which we had obtained species presence/absence information.

Species distribution modelling

We modelled species distributions using an iterative computer learning algorithm called the gradient boosting machine (Friedman 2001; Ridgeway 2006). We used the algorithm as implemented in the package 'gbm: Generalized Boosted Regression Models', available on the R website (http://cran.r-project.org). Boosted regression trees are becoming increasingly popular in niche modelling because of their often superior performance in prediction (Elith *et al.* 2006; Thuiller *et al.* in press). For a full description of gbm and additional references, see Appendix S1 in Supplementary Material. We employed the 'area under the receiver operator characteristic curve' (*AUC*) as a criterion for evaluating the fit of gbm models to the calibration data set and in evaluating predictive performance (Fielding & Bell 1997). This measure of model fit is suited for comparing probabilistic predictions to observed presence–absences because it requires no arbitrarily defined threshold probability with which to establish prediction of species presence. In calibrating the model for each species, we generated an additional estimate of *AUC* by conducting 10-fold cross-validation on a calibration data set that was selected randomly from the complete data set and that conserved the overall proportion of presences/absences that was in the full data set (Efron & Tibshirani 1998; Randin *et al.* 2006). We retrieved directly from the gbm model (in fact, an R-object) the proportional contribution of each climate variable to the model.

We also developed a basis for understanding the statistical significance of differences in model performance among species by bootstrapping 95% percentile confidence intervals for *AUC* values of both model fit (i.e. to the calibration data) and upon prediction (Efron & Tibshirani 1998). To evaluate confidence intervals around the *AUC* value of model fit, for each species we generated 1000 bootstrap data sets from the calibration data (i.e. from the original 1973 observations from the AFE data set). For each bootstrap data set, we fit a gbm model and determined the *AUC* for model fit to the bootstrap data. We established the confidence interval for *AUC* by selecting values corresponding to the upper and lower 2.5% quantiles of the bootstrapped *AUC* distribution. We considered non-overlapping 95% confidence intervals as a sign of significant difference between species *AUC* values. Finally, preparing presence–absence maps for each species from probabilistic predictions requires use of a threshold. Therefore, we used maximized Kappa (Fielding & Bell 1997), choosing the threshold providing the maximum value for the Kappa coefficient of agreement (Cohen 1960), as reported by Randin *et al.* (2006).

Quantifying change in climate–distribution relationships

We determined the relationship between model predictive ability and change in species' estimated niche position by comparing these values between the mid-Holocene and the current period. For both periods, we estimated niche position (i.e. marginality) for all species simultaneously by multiplying the transpose of the species occurrence matrix by the corresponding matrix of normalized observed environmental values. We then conducted a principal components analysis on the difference between estimated current and mid-Holocene niche position for all species (for full details, see Appendix S2). Thus, the change in marginality for each species was calculated relative to the change in marginality of a hypothetical average species in the multivariate space spanned by both current and mid-Holocene climate data sets, using species presence/absence data.

Finally, we assessed model predictive performance as a function of estimated change in niche position by regressing species values of *AUC* on the absolute value of change in niche position from the PCA. We repeated the entire process that is described above to forecast current species distributions based on gbm models that were calibrated with the mid-Holocene data sets.

RESULTS

Species and distributions

We studied seven unambiguously identifiable taxa for which pollen data, when we excluded south-eastern Europe, represented the approximate distributions of the species at the mid-Holocene (Table 1). These species included four trees (*Abies alba*, *Fagus sylvestris*, *Larix decidua* and *Picea abies*) and three shrubs or small trees (*Carpinus betulus*, *Corylus avellana* and *Juniperus communis*) that varied in the AFE data regarding their current distribution across Europe (Figs 1a and 2a; also see Figs S1a–S5a, supplementary online materials). The species varied over an order of magnitude in the degree to which they were represented in pollen count

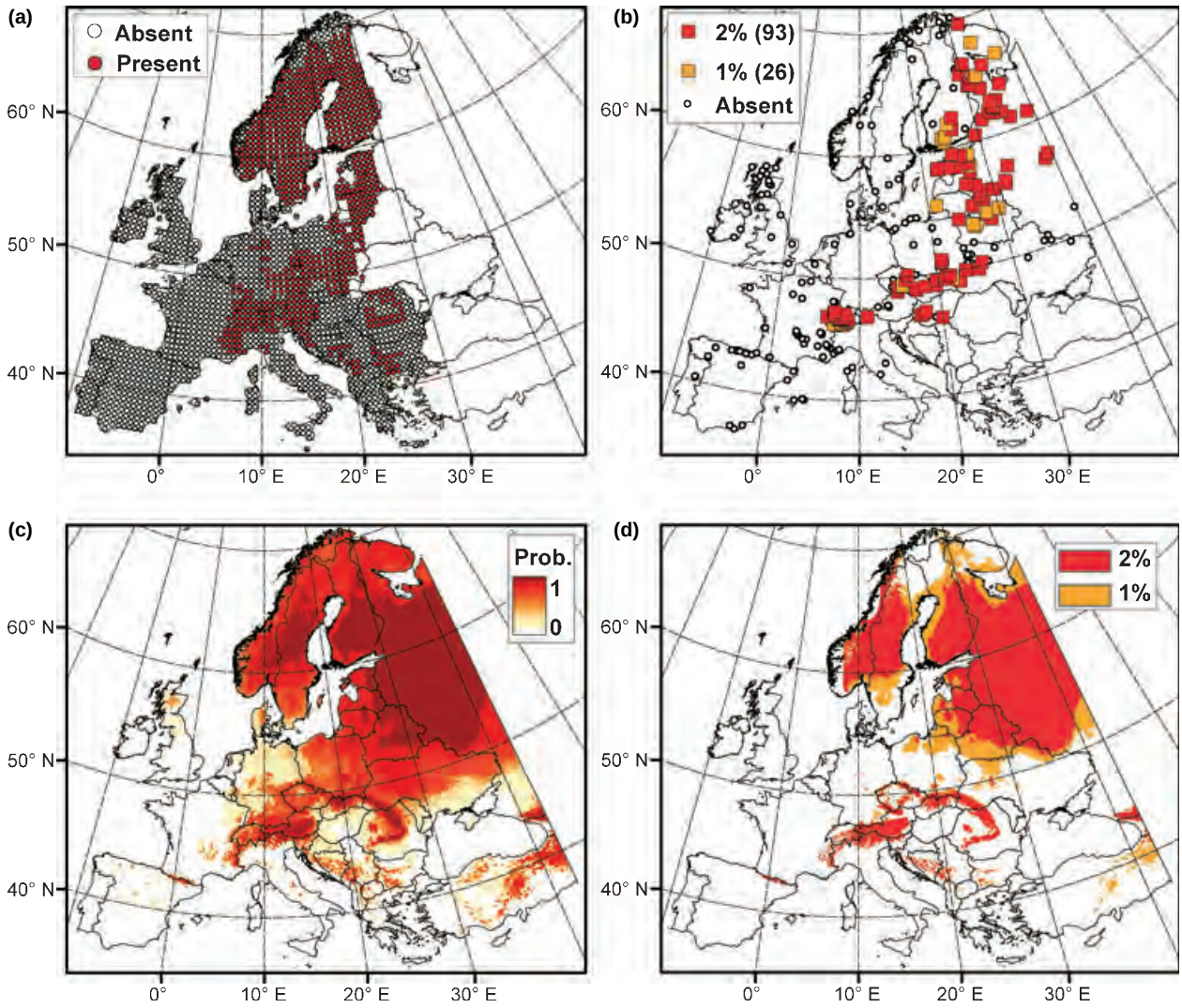


Figure 1 The current and mid-Holocene distributions of *Picea abies*, from empirical data and from niche-based species distribution models (SDMs). (a) The current distribution of the species in Europe, as determined with a digital version of the Atlas Florae Europaeae (AFE). (b) Presence of the species in pollen cores at the mid-Holocene, evaluated at two thresholds. Red squares indicate locations of cores in which pollen frequency of the species reached an average of 2% over the period 6 ky BP (± 500 years). Yellow points indicate additional cores in which the average pollen concentration of the species was ≥ 1% but < 2%. (c) A map of the probability of occurrence of the species currently, obtained by fitting an SDM with AFE data and projecting it over the study area. The probabilities are indicated by a colour band labelled 'Prob.'. (d) Map of the predicted distribution of the species at the mid-Holocene, established using an SDM-generated probability map for the species (i.e. in hindcasting) and a threshold determined by the maximum Kappa criterion, derived from pollen core data. The value of maximum Kappa was calculated using species presence according to high pollen (2%, red) and low pollen (1%, yellow) thresholds. The map thus represents the combination of a model of the estimated current realized niche and pollen data species distribution at 6 ky BP.

data (Table 1, Figs 1b, 2b and S1b–S5b). When evaluated at the high pollen threshold, three species had only a marginally sufficient number of presences (i.e. < 20) for model evaluation on hindcasting and model calibration in forecasting (Table 1).

Modelling results and evaluation

Although models calibrated with current distribution data described well the current distribution of the species (model verification, Figs 1c, 2c and S1c–S5c), model performance was poor for some species upon hindcasting mid-Holocene distributions (model validation; Figs 3a,c and S6a,c). When models were calibrated with current species distribution and climate data, all species models attained an $AUC > 0.8$ and had overlapping 95% confidence intervals for AUC in model verification (Figs 3a and S6a). Thus, we obtained useful models for all species (Swets 1988; Araújo *et al.* 2005) and model performance in fitting the calibration data was statistically indistinguishable among species (Figs 3a and S6). Total annual precipitation together with average annual temperature made by far the strongest contributions to all models (> 90%). In model validation (i.e. evaluation of hindcasted predictions) with the pollen data set and the low threshold for species presence, non-overlapping percentile confidence intervals indicated significant among-species differences in model performance (Fig. 3a), with a notably poor model performance for *J. communis*.

Forecasting current species distributions based on model calibration with pollen-based occurrence data demonstrated substantial uncertainty in the fit of models to the mid-Holocene calibration data, as shown by comparatively wide

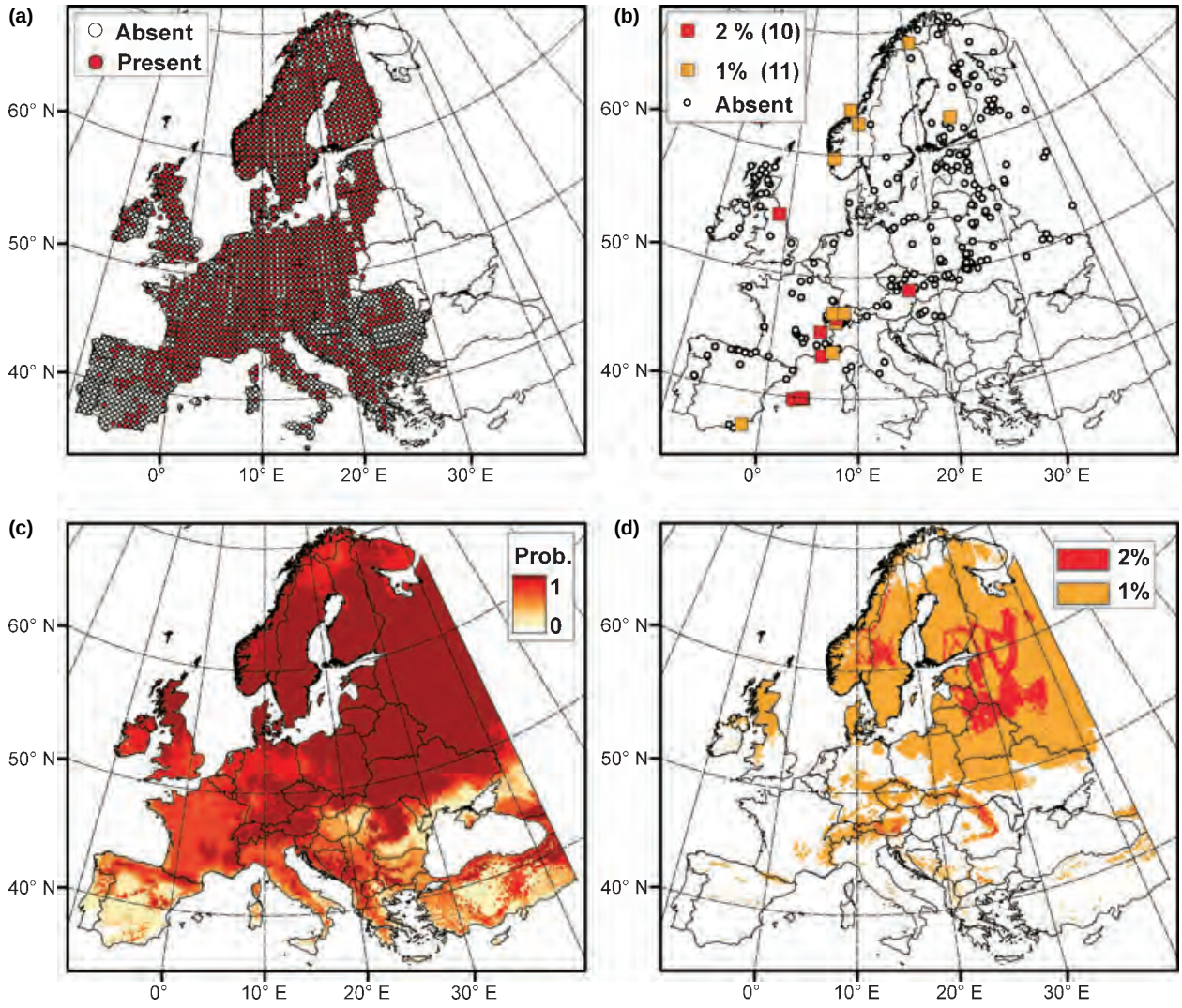


Figure 2 The distribution of *Juniperus communis*, with pollen thresholds as labelled. Otherwise as in Fig. 1.

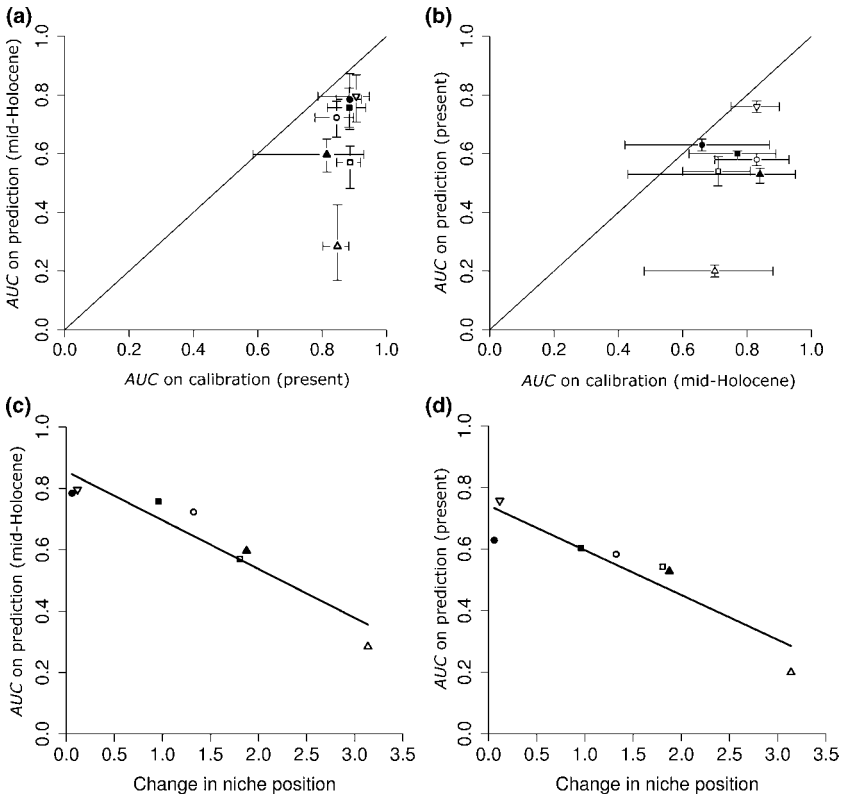


Figure 3 Mean values of the area under the receiver operator characteristic curve (*AUC*) and 95% confidence intervals (a and b), and model performance as a function of magnitude of niche shift (c and d). Top panels show (a) hindcasting with calibration of models using current distribution data and evaluation of predictions based on low pollen percentage thresholds; (b) forecasting current species distributions using mid-Holocene pollen data (low thresholds) for models calibration. Confidence intervals are bootstrapped percentage intervals (95%). Diagonal lines represent equal model performance in fitting and prediction. Species are *Abies alba* (○), *Carpinus betulus* (●), *Corylus avellana* (□), *Fagus sylvatica* (■), *Juniperus communus* (△), *Larix decidua* (▲) and *Picea abies* (▽). The lower panels show the predictive success of models as a function of estimated shift of a species niche position. The *x*-axis is the absolute value of change in estimated niche position, determined in principal component analyses by using low pollen thresholds for species presence. The *y*-axis presents *AUC* values obtained upon prediction, as seen in panels (a) and (b). Evaluations are of hindcasting mid-Holocene distributions (c) and forecasting current species distributions (d). The trend lines are from least-squares regression and are statistically significant. These regressions (c and d) remain significant with removal of *J. communis*. The three colonizing species (*C. avelana*, *J. communus* and *L. decidua*) demonstrate the greatest observed niche shift and the lowest model performance upon prediction.

95% confidence intervals (Figs 3b and S6b). Overlapping confidence intervals indicated that there were no significant differences among species upon model fitting. In contrast, species' models varied greatly in their abilities to forecast current species distributions (Fig. 3b). As in hindcasting, the model for *J. communis* performed much better in describing species occurrence in the calibration data (mid-Holocene) than in prediction (Fig. 3b). Two species' models performed similarly both in fitting the calibration data and in forecasting current distribution (*C. betulus* and *P. abies*). In both the forecasting and hindcasting exercises, there was generally less uncertainty in model performance (i.e. narrower confidence intervals) when using the data sets on current climate and species distributions than when using the mid-Holocene climate estimates and pollen data for model calibration.

Some notable disagreement existed between pollen-based species distribution and hindcasted predictions. *Fagus sylvatica* was predicted to be present in northern Europe but had not occupied this area by 6 ky BP (Fig. S4). *Larix decidua* occurred further south during the mid-Holocene than predicted upon hindcasting (Fig. S5). *Picea abies* was predicted to occupy Scandinavia but there we encountered no pollen evidence for this as of the mid-Holocene (Fig. 1). Finally, *J. communis* was largely absent from central and eastern Europe, although hindcasting predicts that these areas lie within the species potential mid-Holocene range (Fig. 2).

Table 2 Regression statistics and estimated coefficients for two scenarios, in an analysis of model performance (*AUC*) as a function of observed change in species estimated niche positions

Prediction	Pollen threshold	Adjusted r^2	β_0	β_1	SS_{model}	MS_{error}	F*	P-value
Hindcasting	Low	0.87	0.86	−0.16	0.199	0.004	46.97	0.001
Forecasting	Low	0.81	0.74	−0.15	6.42×10^{-14}	2.36×10^{-15}	27.186	0.003

Values for r^2, β_0 and β_1 describe regression lines illustrated in Fig. 1c,d. Statistics were obtained after transformation to meet assumption of normally distributed residuals. See Table S1 in supplementary online material for results using high pollen thresholds.
*All *F*-statistics have one numerator and five denominator degrees of freedom.

Estimated niche position change and model performance

Species varied in estimated difference between their current and mid-Holocene niche positions. *Picea abies* and *C. betulus* experienced less change in estimated niche position when compared to the other species (Figs 3c and S6c). In contrast, the estimated change in niche position of *J. communis* far exceeded that of the other species. The magnitude of species' niche shifts and performance of the models were strongly related, both in hindcasting and in forecasting. Models for species whose niche position had changed little between the mid-Holocene and the present generally performed better than models for species whose niche position had changed substantially relative to the other species (Fig. 3c,d; see also Fig. S6c,d). The significant negative relationships between change in niche position and model performance held for both hindcasting and forecasting, using both low- and high-pollen thresholds. The change in niche position and low *AUC* value of *J. communis* strongly influenced the regression equations (Table 2; see also Table S1). Nonetheless, a significant negative regression persisted when *J. communis* was not included (Fig. 3c,d; analysis not shown).

DISCUSSION

The results demonstrate that the effectiveness of niche-based SDMs in predicting species distributions over a period of 6 ky varies among species. In this way, our results are similar to those on the transfer of SDM predictions in geographic space (Randin *et al.* 2006; Broennimann *et al.* 2007), in which species varied in the accuracy with which they could be predicted in areas other than those where the models were calibrated. These results suggest that work is needed to identify species for which the relationship between distribution and climate likely remains unchanged over relevant time periods. There are a number of reasons why the distributions of some species may be predictable while for others, predictions are markedly erroneous. The realized environmental niche of species may change, either because of changes in either positive or negative biotic interactions in the community or because of evolutionary adaptation to the biotic and abiotic environment (i.e. a change in fundamental niches and realized niches; Broennimann *et al.* 2007; Pearman *et al.* 2008). As change in geographical patterns of climate proceeds, species can also vary in their ability to track these changes. These possibilities suggest further that a change in the mechanisms that limit a species distribution could alter the observed relationship between species distribution and climate (Guisan & Thuiller 2005). Additionally, prediction error might arise from any of the sources of data we used. Thus, understanding which species are more or less likely to be affected by these phenomena will improve confidence in predictions of climate-change impacts provided by niche-based SDMs.

Niche dynamics and species distribution

The degree to which species' niches remain unchanged over time has received substantial attention (Wiens & Graham 2005). Despite this, there has been no clearly identifiable pattern of niche dynamics useful in choosing species that likely provide reliable predictions of climate-change impacts (Pearman *et al.* 2008). We found that species vary in the degree to which estimates of their climate–distribution relationship (i.e. estimated niche position) differed between the mid-Holocene and the present. *Picea abies* and *C. betulus* had little estimated niche change relative to other species. In contrast, the estimated niche and predicted distribution of *J. communis* appears to differ radically between the mid-Holocene and the present. In general, although we have only seven data points, our analysis suggests that species that specialize in disturbed habitats and which are intolerant of shade might have been more likely than competitively dominant species to undergo shifts in estimated climate–distribution relationships during the last 6 ky (Fig. 3).

Likely, at the time scale we investigate here, initial post-glacial expansion of some pioneer species to near their full potential range was followed by reduction in density and/or exclusion from some areas by competitors (Lang 1994). In contrast, we suggest that species could respond differently to rapid climate change (and over a short period), relative to one another, than they do to changes over hundreds or

thousands of years. The actual species for which model projections produce accurate representation of changes in distribution in response to climate change could differ depending on the rate of change and the length of period over which the projections are made. In this regard, competitive ability, time to distribution equilibrium, niche size and the magnitude of displacement of the projection of the niche in geographic space could have large influences. Future research could use hindcasting and other methods to explore these possibilities.

Range filling and dispersal limitation

Similar to cases in which species niches are not stable over time, a tendency for limited range filling for some species would equally suggest that SDM-based predictions of climate-change impact on the geographic distribution of species may be encumbered with error. Incomplete range filling could affect model performance when making predictions of changes in species ranges over time. For example, *L. decidua* might fill < 18% of its current potential range (Svenning & Skov 2004). This species was predicted in hindcasting in the present study to occupy an area extending from the French Alps through the central Alps to eastern Austria. As of the mid-Holocene, evidence for *L. decidua* comes from the central Alps and north of the Adriatic Sea, but is missing from the great portion of the eastern part of its predicted range. A pattern of consistently wide confidence limits for *L. decidua* in both calibration and prediction (Fig. 3a,b) suggests that climate only partially determined the species' climate–distribution relationship by the mid-Holocene. This may in part be due to one or more known ecological properties of *L. decidua*, such as low migration rates, low competitiveness for light, ability to colonize poor soils or anthropogenic effects (Lang 1994; Gobet *et al.* 2003; Bradshaw & Lindbladh 2005). It might also be that although the mid-Holocene climate in central Europe is generally well predicted by GCMs (Brewer *et al.* 2007), some errors in climate estimation could be influential. Comparison of models derived from a variety of GCMs could be informative regarding these effects.

In another example, *F. sylvatica* was missing from much of the northerly portion of its predicted range as of the mid-Holocene (Fig. S4). Our models over-predict the mid-Holocene range of the species in northern Europe, as do physiological models (Giesecke *et al.* 2007). Those models predict a much broader distribution as of 6 ky BP than do the SDM results described here, although the two results are not directly comparable because of differing methodology and response variables. *Fagus sylvatica* only spread into its predicted suitable range with the advent of anthropogenic fires in the late-Holocene, in areas that up to that point were dominated by *P. abies* (Gobet *et al.* 2003; Bradshaw & Lindbladh 2005). We cannot, using only our data, distinguish with certainty the degree to which observed changes in estimated niche position and geographic range filling result from niche shift, competitive exclusion or some degree of dispersal limitation. Nonetheless, these results suggest that the mechanisms limiting the distribution of *F. sylvatica* may have changed over 6 ky.

Mechanisms influencing species distribution and niche position

Multiple processes surely influence species distributions and might change in importance over time, to the degree that the role of climate in limiting species distributions could have changed between the mid-Holocene and the present. This could result in a change in estimated niche position in multivariate climate space. For example, *P. abies* was a widely distributed species as of the mid-Holocene (Fig. 1b). Notably, *J. communis* had a similarly large predicted mid-Holocene range as *P. abies*. Pollen records show that *J. communis* was common and present at some sites in northern and western Europe before arrival of *P. abies* and full forest development (Lang 1994), suggesting that the late glacial and early Holocene distribution of *J. communis* was limited primarily by climate, not by competition. However, little pollen evidence for *J. communis* presence is available from much of its predicted mid-Holocene range in central Europe (Fig. 2b). Much of this area was by this time occupied by *P. abies* and mixed forest species (Lang 1994). *Juniperus communis* is a colonizer of disturbed and open areas due to its requirement for light and can tolerate soils with little development (Zoller 1981). This suggests that as of the mid-Holocene, competition had a larger influence on the distribution of *J. communis* than did climate.

It is unlikely that *J. communis* was completely absent from most of central Europe as of the mid-Holocene (i.e. absence of evidence is not evidence of absence), but the patterns we observe suggest that this species was at low density or absent across extensive areas of its potential range, which had come to be dominated by *P. abies* and other trees of closed canopy forest. The broad range but scattered distribution of *J. communis* in recent times (Fig. 2a) suggests that the species has benefited greatly from expanding anthropogenic disturbance, but may still be excluded from many optimal habitats by stronger competitors. Similar reasoning may explain the relatively poor fit of hindcasted models of *C. avellana*, a species which may have been absent locally within its predicted mid-Holocene range because of competition for light in closed canopy forest (Oberdorfer 1990; Burga & Perret 1998).

One further cause of prediction errors is that the mechanisms limiting species distributions could be hetero-

geneous, not just in time, but also over the range of the species. For example, it might be that in mapping *J. communis* occurrences in the mid-Holocene we included more (nominally) intraspecific variation by pooling closely related species or ecologically distinct populations when making or evaluating predictions. Another possibility is that just one or a few populations could be responsible for changes in species distribution. A case of differential range (and niche?) expansion of populations is known for *F. sylvatica*, in which only one population is thought to have expanded from its Pleistocene refugium to cover much of western Europe (Magri *et al.* 2006). This suggests that a niche shift affected this population, perhaps similar to the expansion of populations of invasive species in their introduced range (Broennimann *et al.* 2007). Intraspecific variation may also influence the local distribution and degree to which *L. decidua* is excluded by forest species such as *A. alba* and *F. sylvatica* (Zoller 1981).

Paleo-data, SDMs and prediction uncertainty

Fossil pollen data is often incomplete and the adequacy of the pollen record in representing species prior distributions can vary among species. For example, *L. decidua* produces and spreads small amounts of pollen (Zoller 1981). The pollen data on *L. decidua*, even at the low threshold we used here, probably underestimate the distribution of the tree. *Larix decidua* grows densely in a narrow elevation belt in mountains today (Fig. S5). Thus, pollen sites outside this belt likely record very little or no *L. decidua* pollen even when the species occurs relatively short distances away.

In another example, *P. abies* potentially might fill < 85% of its potential range currently (Svenning & Skov 2004). While the authors of that study (as here) may have failed to include one or more variables that are important to determining species distribution, this does not appear to affect model performance greatly upon hindcasting the mid-Holocene distribution of this species. Other species may be affected by choice of climate variables, but this is not possible to evaluate here. One possible explanation is that SDM performance may depend more on the degree to which a species has expanded to the limits of its niche than on expansion to its potential geographic range. Another possibility is that the spatial distribution of pollen cores might influence model performance differentially among species. For example, *P. abies* was absent from areas in Scandinavia that we predicted as suitable habitat as of the mid-Holocene. Over-prediction of the species in this area could conceivably have little influence on model fit because of the paucity in our data set of pollen cores from the area that is over-predicted, which corresponds to central Sweden (Fig. 1).

Limitations to the approach and future directions

We have shown that hindcasting and forecasting of species distributions using data from pollen cores are ways to evaluate model performance across time. While we used all available pollen data we could acquire electronically, continued growth of pollen databases will allow improved assessment of the effects on model performance of the choice of pollen thresholds. Increasing availability of plant macrofossils may eventually be sufficient as an additional source of data for modelling the distributions of some species. This may improve projections for species that can be present at low densities or otherwise produce little pollen.

A further limitation is that the equilibrium distribution of species cannot be estimated independently from competitive effects. Ideally, one would estimate the fundamental niche of a species, then constrain it with estimates of the effects of competitive interactions (Guisan *et al.* 2007b; Pearman *et al.* 2008). However, even with an understanding of the physiological limits that determine the fundamental niche, an estimation of the equilibrium distribution will require an operational understanding of how competitive interactions constrain the fundamental niche and result in an expected distribution based on the realized niche.

Future research on predictive modelling may result in better techniques, or an ensemble of techniques, with which to address effects of migration rates and predict levels of range filling (Botkin *et al.* 2007; Thuiller *et al.* in press). In the present study, we faced the additional problem of choosing a GCM for estimating past climate, because we were further constrained not to use comparison of GCM results with pollen-based climate reconstructions as a basis for our choice. Future work may approach effects of variation in predictions of past climate as a random factor to quantitatively estimate uncertainty generated by choice of climate model. For example, by model performance could be evaluated with a larger number of historical GCM runs for mid-Holocene climate (e.g. Brewer *et al.* 2007). Furthermore, variability in data quality that arises from any of our data sources could influence the observed relationship between distribution and climate (i.e. the realized climatic niche). Consideration of these issues and careful choice of additional species for analysis, both in Europe and elsewhere, will help improve our understanding of the ecological characteristics of species and the structural characteristics of data sets that influence predictions of plant response to climate change.

CONCLUSIONS

The increasing availability of data on plant distribution and climate allows testing SDMs by hindcasting and

forecasting. Model performance upon prediction depends on the estimated shift in a species' climate–distribution relationship. The ability to compete for light may be one factor that influences the relationship between a species distribution and geographic variability in climate. Additional effects likely include those of dispersal (migration) ability and range filling, elapsed time over which a prediction is made, and qualities of invariably imperfect data. The long time span used here, relative to predicted future climate change, likely affected in contrasting ways the estimated niche position changes in the study species. Future research should seek to improve quantitative understanding of the effects of these factors on predictions across time that are made with SDMs.

ACKNOWLEDGEMENTS

We thank C. Calenge for discussion of niche shift and related multivariate analysis and we acknowledge the constructive comments of two anonymous referees on a previous version of this paper. We acknowledge the international modelling groups for providing their data for analysis, and the Laboratoire des Sciences du Climat et de l'Environnement (LSCE) for collecting and archiving the GCM model data. The PMIP2/MOTIF Data Archive is supported by CEA, CNRS, the EU project MOTIF (EVK2-CT-2002–00153) and the Programme National d'Etude de la Dynamique du Climat (PNEDC). The analyses were performed using data downloaded on 31 March 2006. More information is available on http://www-lsce.cea.fr/pmip2/. Pollen data were in part extracted from the European Pollen Database http://www.europeanpollendatabase.net and the Alpine Palynological Database (ALPADABA, University of Bern, Switzerland). This study was supported by the Swiss National Science Foundation (grant no. 110000) and the European Science Foundation (FP6 ECOCHANGE and MACIS projects).

REFERENCES

Araújo, M.B. & Rahbek, C. (2006). How does climate change affect biodiversity? *Science*, 313, 1396–1397.

Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005). Validation of species-climate impact models under climate change. *Glob. Chang. Biol.*, 11, 1504–1513.

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.

Botkin, D.B., Saxe, H., Araujo, M.B., Betts, R., Bradshaw, R.H.W., Cedhagen, T. *et al.* (2007). Forecasting the effects of global warming on biodiversity. *Bioscience*, 57, 227–236.

Bradshaw, R.H.W. & Lindbladh, M. (2005). Regional spread and stand-scale establishment of *Fagus sylvatica* and *Picea abies* in Scandinavia. *Ecology*, 86, 1679–1686.

Brewer, S., Guiot, J. & Torre, F. (2007). Mid-Holocene climate change in Europe: a data-model comparison. *Clim. Past*, 3, 499–512.

Broennimann, O., Thuiller, W., Hughes, G., Midgley, G.F., Alkemade, J.R.M. & Guisan, A. (2006). Do geographic distribution, niche property and life form explain plants' vulnerability to global change? *Global Change Biology*, 12, 1079–1093.

Broennimann, O., Treier, U., Müller-Schärer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007). Evidence of climatic niche shift during biological invasion. *Ecol. Lett.*, 10, 701–709.

Burga, C.A. & Perret, R. (1998). *Vegetation und Klima der Schweiz seit dem jüngeren Eiszeitalter*. Ott Verlag, Thun, Switzerland.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, 20, 37–46.

Davis, M.B. & Shaw, R.G. (2001). Range shifts and adaptive responses to Quaternary climate change. *Science*, 292, 673–679.

Davis, B.A.S., Brewer, S., Stevenson, A.C. & Guiot, J. (2003). The temperature of Europe during the Holocene reconstructed from pollen data. *Quatern. Sci. Rev.*, 22, 1701–1716.

Dolédec, S., Chessel, D. & Gimaret-Carpentier, C. (2000). Niche separation in community analysis: a new method. *Ecology*, 81, 2914–2927.

Efron, B. & Tibshirani, R.J. (1998). *An Introduction to the Bootstrap*. Chapman and Hall/CRC Press, Boca Raton, Florida.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A. *et al.* (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129–151.

ESRI (2005). *ARCInfo Version 9.1*. Environmental Systems Research Institute, Redlands, California.

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.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.*, 29, 1189–1232.

Gehrig-Fasel, J., Guisan, A. & Zimmermann, N.E. (2007). Tree line shifts in the Swiss Alps: climate change or land abandonment? *J. Veg. Sci.*, 18, 571–582.

Giesecke, T., Hickler, T., Kunkel, T., Sykes, M.T. & Bradshaw, R.H.W. (2007). Towards an understanding of the Holocene distribution of *Fagus sylvatica,* L. *J. Biogeogr.*, 34, 118–131.

Gladstone, R.M., Ross, I., Valdes, P.J., Abe-Ouchi, A., Braconnot, P., Brewer, S. *et al.* (2005). Mid-Holocene NAO: a PMIP2 model intercomparison. *Geophys. Res. Lett.*, 32 (16), Article No. L16707 2005, doi: 10.1029/2005GL023596.

Gobet, E., Tinner, W., Hochuli, P.A., van Leeuwen, J.F.N. & Ammann, B. (2003). Middle to late Holocene vegetation history of the Upper Engadine (Swiss Alps): the role of man and fire. *Veg. Hist. Archaeobot.*, 12, 143–163.

Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C. *et al.* (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.*, 16, 147–168.

Grabherr, G., Gottfried, M. & Pauli, H. (1994). Climate effects on mountain plants. *Nature*, 369, 448.

Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.*, 8, 993–1009.

Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecol. Modell.*, 135, 147–186.

Guisan, A., Graham, C.H., Elith, J. & Huettmann, F. (2007a). Sensitivity of predictive species distribution models to change in grain size. *Divers. Distrib.*, 13, 332–340.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. (2007b). What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecol. Monogr.*, 77, 615–630.

Hannah, L., Midgley, G.F. & Millar, D. (2002). Climate change-integrated conservation strategies. *Glob. Ecol. Biogeogr.*, 11, 485–495.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.*, 25, 1965–1978.

Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Dai, X. *et al.* (2001). *Climate Change 2001: The Scientific Basis. Contributions of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge.

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. *J. Biogeogr.*, 16, 551–560.

Hutchinson, G.E. (1957). Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.*, 22, 415–427.

Jalas, J. & Suominen, J. (1972–1999). *Atlas Florae Europaeae: Distribution of Vascular Plants in Europe*. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.

Kullman, L. (2001). Immigration of *Picea abies* into North-Central Sweden. New evidence of regional expansion and tree-limit evolution. *Nord. J. Bot.*, 21, 39–54.

Lang, G. (1994). *Quartäre Vegetationsgeschichte Europas. Methoden und Ergebnisse*. Gustav Fischer Verlag, Jena, Germany.

Latalowa, M. & van der Knaap, W.O. (2006). Late Quarternary expansion of Norway spruce *Picea alba* (L.) Karst. in Europe according to pollen data. *J. Quaternary Sci.*, 25, 2780–2805.

Magri, D., Vendramin, G.G., Comps, B., Dupanloup, I., Geburek, T., Gomory, D. *et al.* (2006). A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytol.*, 171, 199–221.

Martinez-Meyer, E. & Peterson, A.T. (2006). Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *J. Biogeogr.*, 33, 1779–1789.

Oberdorfer, E. (1990). *Pflanzensoziologische Exkursionsflora*, 6th edn. Ulmer, Stuttgart.

Parmesan, C. & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37–42.

Pearman, P.B., Guisan, A., Broennimann, O. & Randin, C.F. (2008). Niche dynamics in space and time. *Trends Ecol. Evol.*, (in press).

Pearson, R.G. & Dawson, T.P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Glob. Ecol. Biogeogr.*, 12, 361–371.

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.

Randin, C.F., Dirnbock, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006). Are niche-based species distribution models transferable in space? *J. Biogeogr.*, 33, 1689–1703.

Ridgeway, G. (2006). *Generalized Boosted Models: A Guide to the gbm Package. R vignette. [unpublished PDF document]*. Available at: http://www.i-pensieri.com/gregr/papers/gbm-vignette.pdf. Last accessed on 2 January 2008.

Svenning, J.C. & Skov, F. (2004). Limited filling of the potential range in European tree species. *Ecol. Lett.*, 7, 565–573.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C. *et al.* (2004). Extinction risk from climate change. *Nature*, 427, 145–148.

Thuiller, W., Lavorel, S., Araujo, M.B., Sykes, M.T. & Prentice, I.C. (2005). Climate change threats to plant diversity in Europe. *Proc. Natl Acad. Sci. U.S.A.*, 102, 8245–8250.

Thuiller, W., Albert, C., Araujo, M.B., Berry, P.M., Cabeza, M., Guisan, A. *et al.* (in press). Predicting global change impacts on plant species' distributions: future challenges. *Perspect. Plant Ecol. Evol. Syst.*

Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C. *et al.* (2002). Ecological responses to recent climate change. *Nature*, 416, 389–395.

Wiens, J.J. & Graham, C.H. (2005). Niche conservatism: integrating evolution, ecology, and conservation biology. *Annu. Rev. Ecol., Evol. Syst.*, 36, 519–539.

Zoller, H. (1981). Gymnospermae. In: *Illustrierte Flora von Mitteleuropa* (eds Conert, H.J., Hamann, U., Schultze-Motel, W. & Wagenitz, G.). Paul Parea, Berlin, pp. 11–148.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Appendix S1 Niche and distribution modeling: boosted regression trees.

Appendix S2 Determination of change in estimated niche position in PCA climate space between two time periods.

Figure S1 The distribution of *Abies alba*, with pollen thresholds as labeled.

Figure S2 The distribution of *Carpinus betula*, with pollen thresholds as labeled.

Figure S3 The distribution of *Corylus avellana*, with pollen thresholds as labeled.

Figure S4 The distribution of *Fagus sylvatica*, with pollen thresholds as labeled.

Figure S5 The distribution of *Larix decidua*, with pollen thresholds as labeled.

Figure S6 Mean values of the area under the receiver operator characteristic curve (AUC) and 95% confidence intervals (a, b), and model performance as a function of magnitude of shift in climate space (c, d).

Table S1 Regression statistics and estimates of coefficients for two scenarios, in an analysis of model predictive ability (AUC) as a function of observed change in species realized climate niche position.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/full/10.1111/j.1461-0248.2007.01150.x.

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Editor, John Harte
Manuscript received 20 September 2007
First decision made 31 October 2007
Manuscript accepted 6 December 2007

Appendix S1 Niche and distribution modeling: boosted regression trees

Examples of boosted regression trees that are available in the ecological literature potentially leave the impression that *gbm* is a black box method. In the *gbm* algorithm (Friedman 2001, 2002), model fitting occurs not in parameter space but instead in function space. The *gbm* iteratively fits shallow regression trees, updating a base function with additional regression tree models. In contrast to earlier boosting algorithms that re-weight poorly fitted observations with each iteration, the *gbm* incorporates a 'negative gradient' criterion (Ridgeway 1999, 2005). Under this criterion, the 'loss' function (i.e., residual error, e.g., sum of mean squared residuals) is evaluated at each iteration of the algorithm to determine what new function will minimize the loss function's resulting expected value (Friedman 2002; Ridgeway 2005).

We designated that a randomly chosen half of the training data be used for function fitting, leaving the other half for estimating the optimal number of trees to use during prediction with the model (out-of-bag estimate, below). When splitting the dataset in half, we maintained the relative prevalence of species. Further, we specified that with each iteration, an element of randomness be established by picking randomly without replacement 50% of the training data to generate the new fitted function (i.e., 'bagging' (Breiman 1996)). Thus, 25% of the current data were used for each estimate of the new function with which to update the existing function. The new function was then multiplied by 0.001 (the 'shrinkage value') to decrease the size of the step towards the minimized expected value of the loss function, effectively slowing the algorithm's per iteration learning rate. Bagging during function fitting and application of small shrinkage values increase the predictive ability of the calibrated model object (Ridgeway 2005). The gbm algorithm updated (by addition) the base function from the

previous iteration with the new shrunk function. We used two thousand iterations to produce final *gbm* objects, each with 2000 trees.

To predict plant species distribution using the resulting *gbm* model object, we chose the 'oob' (out-of-bag) criterion (Breiman 1996) to estimate what fraction of the trees in the final *gbm* model object should be used to form the optimal predictive function. With the oob criterion, the estimated prediction error (deviance) after each subsequent tree was evaluated by predicting on the random fraction of the data that were not used in constructing the subsequent regression tree. This criterion is conservative and results in a final predicting function, the performance of which is not compromised by over-fitting of the training data (Ridgeway 2005). Then, using the final function developed for each species, we hindcasted probabilities of species occurrence in the study area at 6 ky BP using climate data from the general circulation model. We similarly fit gbm objects using species presence/absence determined from pollen core data and GCM-estimated climate to forecast current species distributions.

Supplementary Literature Cited

Breiman L. (1996) Bagging predictors. *Machine Learning*, 24, 123-140

Friedman J.H. (2001) Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29, 1189-1232

Friedman J.H. (2002) Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38, 367-378

Ridgeway G. (1999) The state of boosting. *Computing Science and Statistics*, 31, 172-181

Ridgeway G. (2005) Generalized Boosted Models: A guide to the gbm package [PDF document]. URL http://www.i-pensieri.com/gregr/papers/gbm-vignette.pdf

Appendix S2 Determination of change in estimated niche position in PCA climate space between two time periods

Let *e* be a number of environmental variables and let **E** be a partitioned matrix consisting of two submatrices

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix},$$

where $\mathbf{E}_1$ and $\mathbf{E}_2$ have dimensions (m_1 x e) and [m_2 x e] respectively, m_x indicating the number of observations (rows) in each partition. These submatrices correspond to environmental measurements taken at the two different times.

We calculated the matrix $\mathbf{E}^*$, in which each element of **E** has been standardized by subtracting the corresponding column mean of **E** and dividing by the corresponding column's standard deviation, to form:

$$\mathbf{E}^* = \begin{bmatrix} \mathbf{E}_1^* \\ \mathbf{E}_2^* \end{bmatrix} .$$

This centered each submatrix's values over a single, inclusive environmental space.

We then calculated matrices of 'marginality' vectors for each species (with reference to the center of observed environmental space) using the two matrices whose elements were the presence (1) or absences (0) of the *s* species, each matrix of marginality vectors corresponding to the submatrices of standardized environmental variables,

$$\mathbf{X}_1 = \mathbf{S}_1^{\mathrm{T}} \mathbf{E}_1^*$$
$$\mathbf{X}_2 = \mathbf{S}_2^{\mathrm{T}} \mathbf{E}_2^*$$

where $\mathbf{S}_1$ and $\mathbf{S}_2$ have dimensions (m_1 x s) and (m_2 x s).

We then calculated the difference in marginality between the two time periods as

$$\mathbf{X} = \mathbf{X}_1 - \mathbf{X}_2 ,$$

and performed a non-centered principal component analysis on **X**. This maximized the mean difference between centroids for the species. We extracted the vector **u** of eigenvalues of **X**, and pre-multiplied it by $\mathbf{X}_1$ and $\mathbf{X}_2$, and then took the absolute values in order to calculate the absolute magnitude of observed niche position change ($\Delta\mathbf{p}$) for each species,

$$\left|\Delta\mathbf{p}\right| = \left|\mathbf{X}_1\mathbf{u} - \mathbf{X}_2\mathbf{u}\right| .$$

Figure S1 The distribution of *Abies alba*, with pollen thresholds as labeled. Otherwise as in Figure 1.

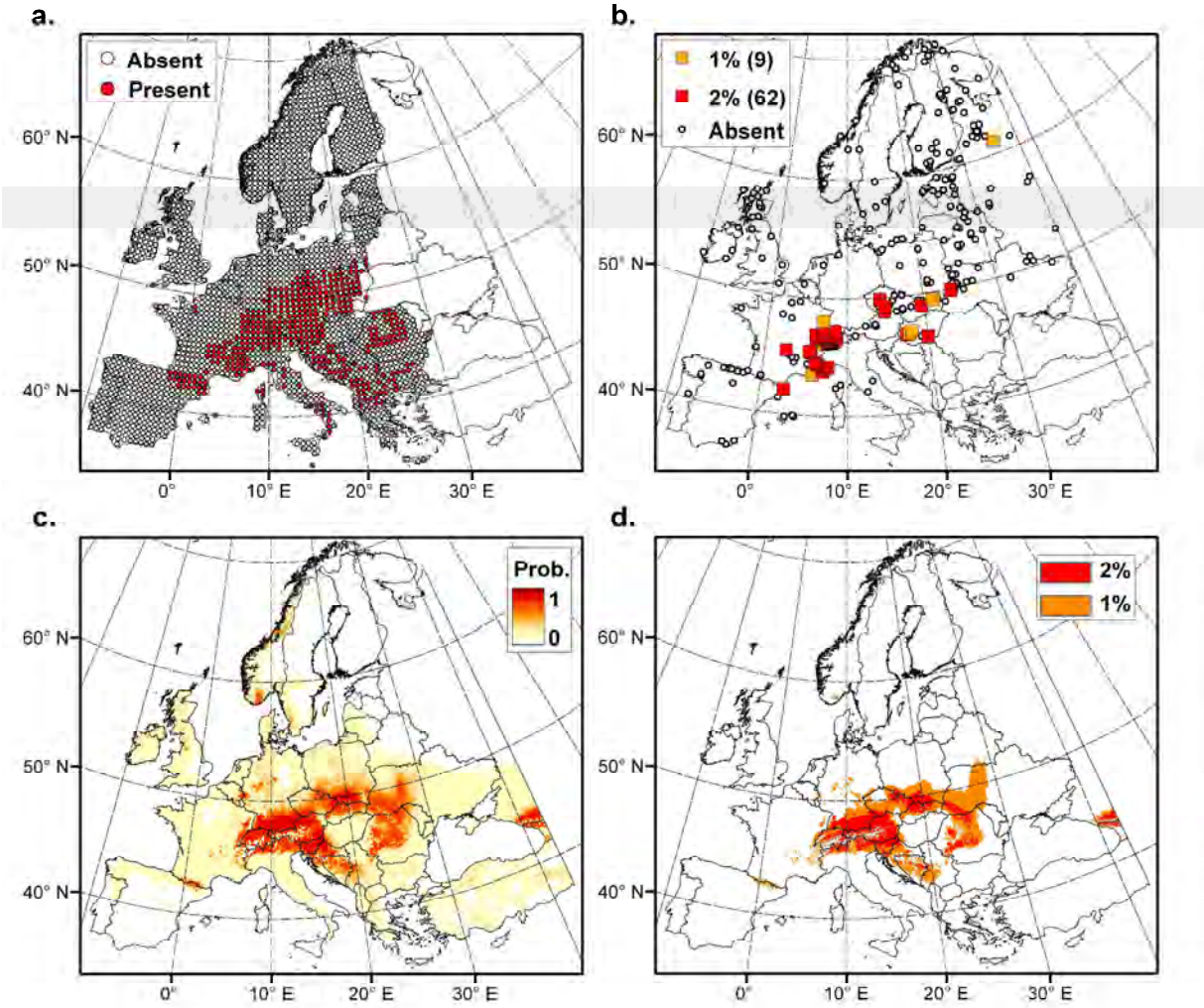


Figure S2 The distribution of *Carpinus betula*, with pollen thresholds as labeled. Otherwise as in Figure 1.

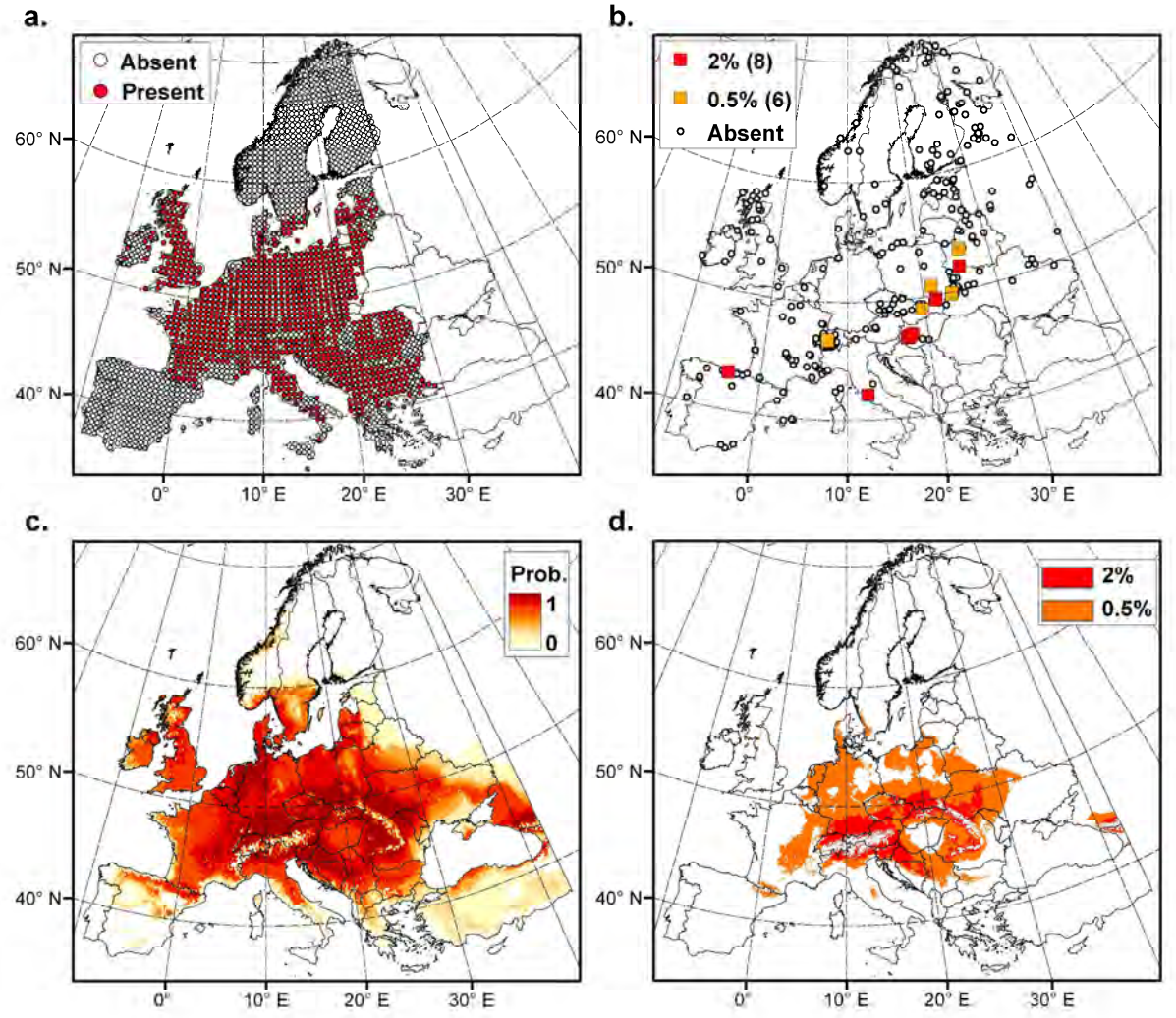


Figure S3 The distribution of *Corylus avellana*, with pollen thresholds as labeled. Otherwise as in Figure 1.

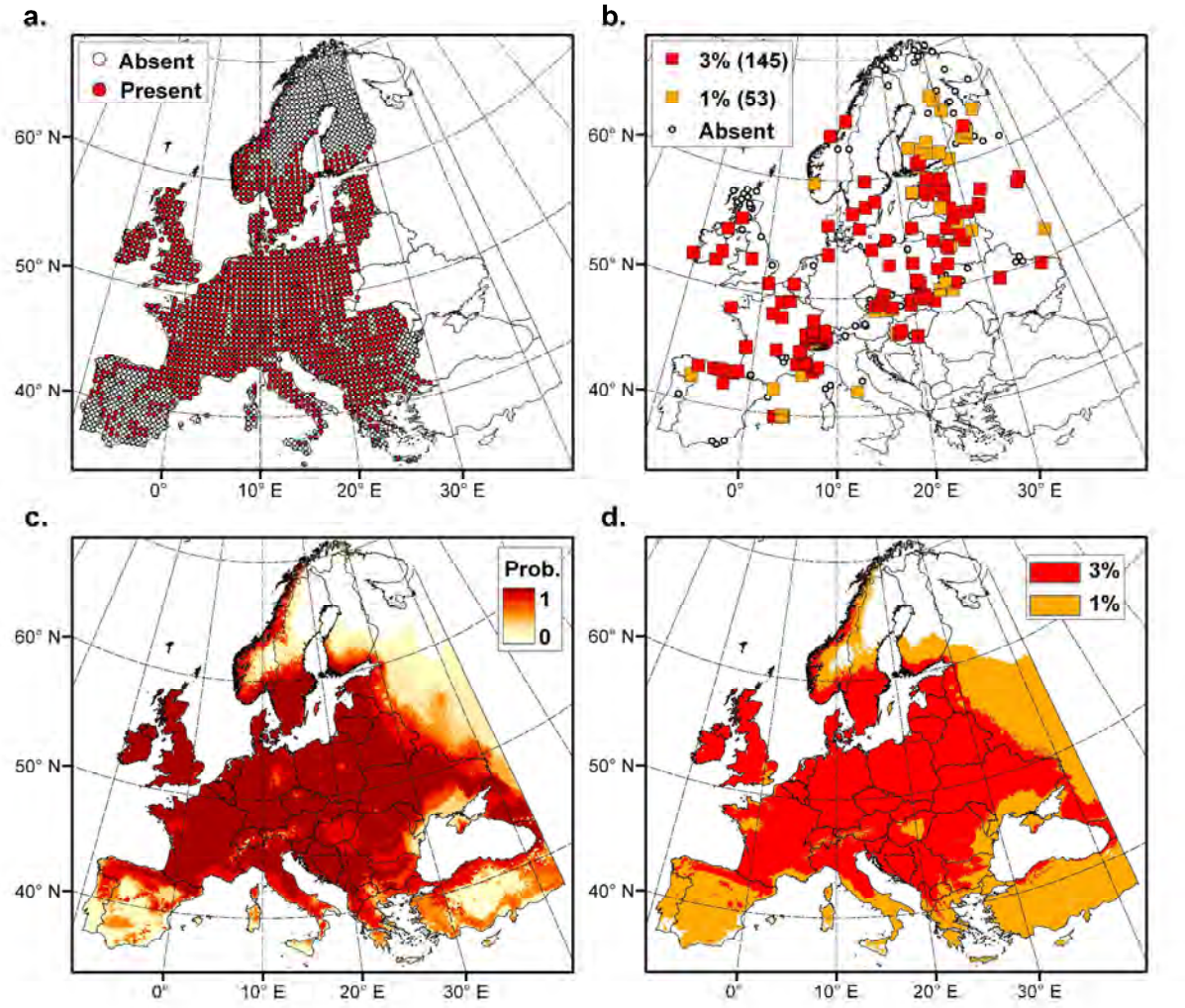


Figure S4 The distribution of *Fagus sylvatica*, with pollen thresholds as labeled. Otherwise as in Figure 1.

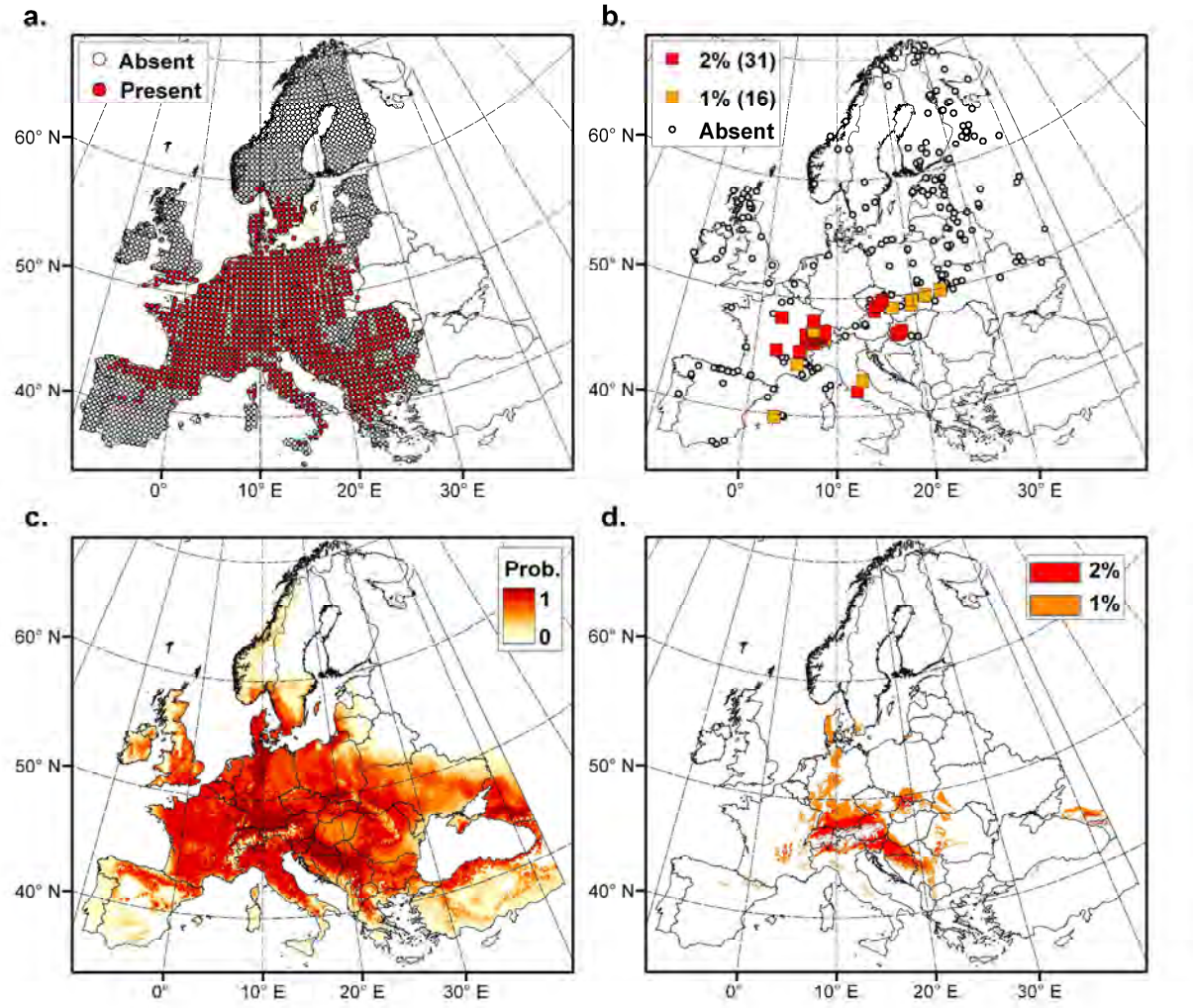


Figure S5 The distribution of *Larix decidua*, with pollen thresholds as labeled.

Otherwise as in Figure 1.

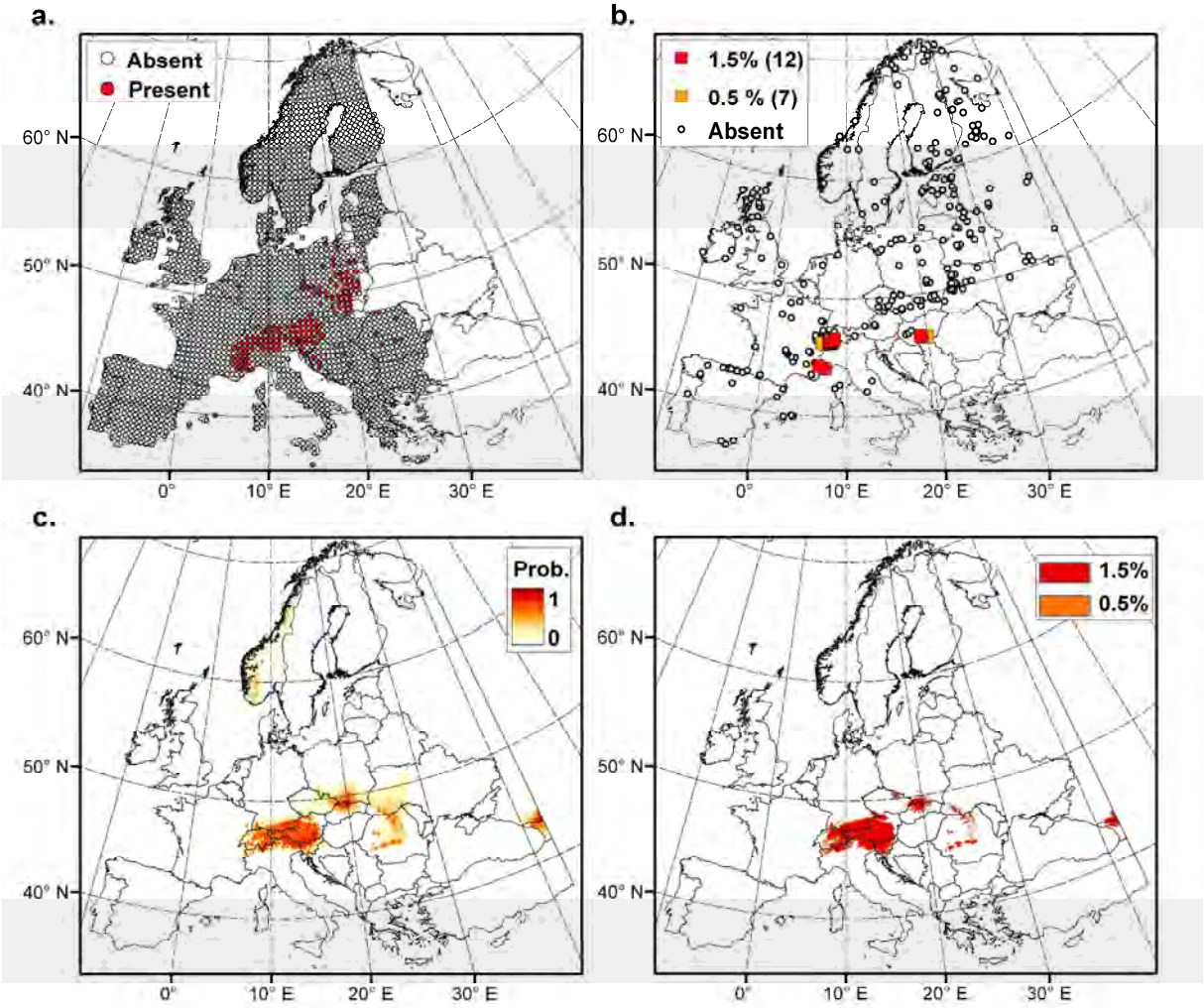


Figure S6 Mean values of the area under the receiver operator characteristic curve (AUC) and 95% confidence intervals (a, b), and model performance as a function of magnitude of shift in climate space (c, d). As in Figure 3, except that analyses use the high pollen threshold for species presence at the mid-Holocene.

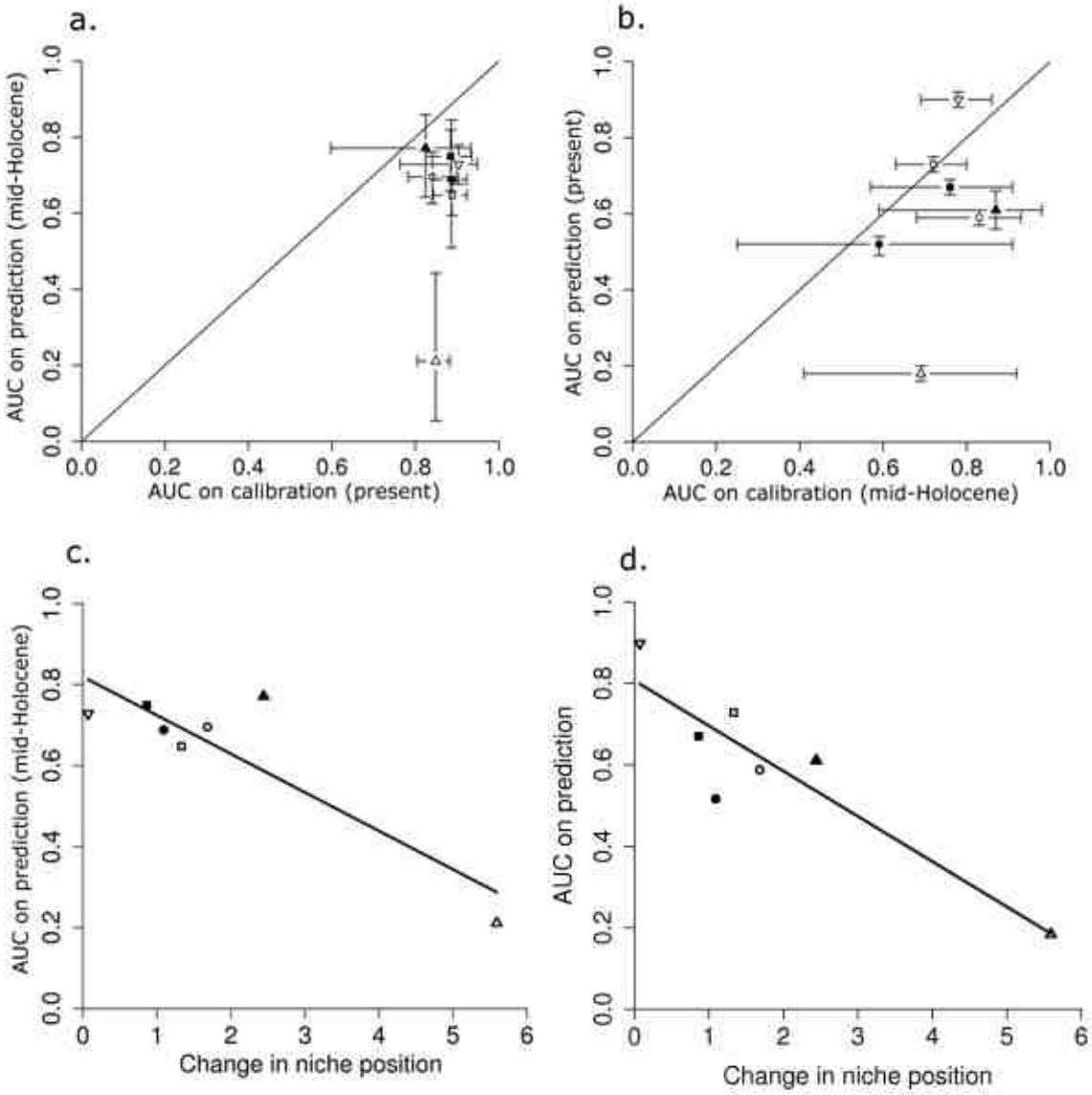


Table S1 Regression statistics and estimates of coefficients for two scenarios, in an analysis of model predictive ability (AUC) as a function of observed change in species realized climate niche position.

Values for r^2, β_0 and β_1 describe regression lines shown in Figure S6c, d. Statistical values were obtained after transformation to meet assumption of normally distributed residuals. Species presence at mid-Holocene was evaluated with high pollen thresholds.

Prediction	**Pollen thresholds**	**adj. r^2**	**β_0**	**β_1**	**SS_{model}**	**MS_{error}**	**F***	**P**
Hindcasting	High	0.73	0.82	-0.095	20341.5	746.8	27.24	0.003
Forecasting	High	0.79	0.81	-0.11	0.239	0.010	23.10	0.005

* All F-statistics have one numerator and 5 denominator degrees of freedom.

A 5

MIGCLIM User Guide.

Robin Engler

MigClim User Guide

by Robin Engler, August 2009.

Table of Contents

About this user guide

This is version 14-08-2009 of the MigClim user guide.

Part 1: MigClim overview.

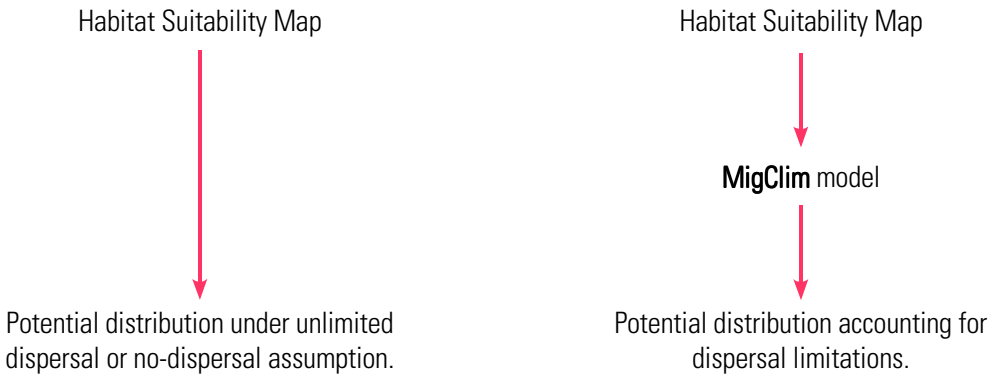
MigClim quick overview

When using "species distribution models" (also known as "habitat suitability models" or "Resource selection functions") to identify the potential distribution of a species, one recurrent problem is to account for species dispersal limitations.

For instance, if we project the potential distribution of a species into the future under a climate change scenario, can we be sure that our species will really occupy all of its potential habitats? In other words, this asks the questions whether the assumption of unlimited dispersal a realistic one. There are potentially many elements that challenge this assumption, and impede a species to reach all of its potential habitat. For instance, the increasing fragmentation of the landscape could be a barrier to species dispersal, or at least slow down its dispersal rate.

Similarly if one attempts to model an invasive species' potential distribution, the dispersal limitation could also be an important parameter to take into account. More specifically, in the case of an invasive species, the most important question is maybe not whether or not all potentially suitable locations will eventually become colonized (generally invasive species are rather good at dispersing), but when they might become colonized and which is the most likely route that the species will take to spread through the landscape. Again, accounting for dispersal appears as an important issue in this context.

The MigClim model has been developed to address this kind of issues. It allows accounting for dispersal limitations when making projections of potential species distribution. It can be thought-off as a supplementary component of a species distribution model, that acts between the habitat suitability predictions and the final potential distribution of a species, by restricting this potential distribution through dispersal limitations.



MigClim is an add-on for the ArcGIS® software (www.esri.com) primarily designed to simulate dispersal of plant species in fragmented landscapes and under environmental change scenarios (e.g. climate change scenarios). It is obviously also possible to model dispersal without implementing environmental change into the simulation, for instance if one wishes to model the potential spread of an invasive species.

Although MigClim was primarily developed to model dispersal of plant species, it is also possible, bearing it's limitations in mind, to apply it to animal species.

Because the idea behind MigClim is to provide users with a relatively easy to calibrate and flexible model that can adapt to many species, MigClim does currently not implement any advanced population dynamics parameters such as number of individuals per pixel or number of seeds produced per individual. MigClim is also limited in the number of pixels it can handle. The maximum number of pixels

supported is of about 50 millions, but it can vary depending on how many parameters are implemented in a given simulation (and possibly also on the power of the computer used to run the simulations).

To get a better idea of why the MigClim model was developed, how it works and how it can be applied to many species, I recommend reading the two following (highly entertaining ;-) scientific publications:

→ Engler R. and Guisan A., 2009. MigClim: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, **15** (4), 590-601.

→ Engler R., Randin C.F., Vittoz P., Czáka T., Beniston M., Zimmermann N.E. and Guisan A, 2009. Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? *Ecography*, **32** (1), 34-45.

Engler and Guisan (2009) introduces the MigClim model, explains its purpose and how it works, and gives an example of application for two semi-virtual species. Engler *et al.* (2009) provides an example of the application of MigClim to a large number of species and could thus give you some indications on how to calibrate the model for a large number of species with limited available data.

How to cite the MigClim model

If you use MigClim, please cite it as follows:

> Engler R. and Guisan A., 2009. MigClim: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, **15** (4), 590-601.

Parameters currently implemented in MigClim

The parameters that are currently implemented in the MigClim model are the following:

- Environmental change over time (e.g. climate change scenario).
- Probability of pixel colonization given its distance to a source cell.
- Long distance dispersal (LDD).
- Increase of a pixel's (representing a population) reproductive potential over time. This parameter also includes time for a pixel to reach reproductive maturity.
- Barriers to dispersal.
- Permanent unsuitable locations.
- Random population extinctions.
- Vegetative resilience time.
- Seed bank resilience time.
- Differential dispersal along river features.
- Differential dispersal along road features.
- Long distance dispersal along river features.
- Long distance dispersal along road features.
- Anisotropic dispersal by wind.
- Anisotropic dispersal by terrain slope.

Important: The basic modeling unit of MigClim is a pixel, representing a population. Therefore, all parameters values refer to population values (i.e. not values for individual plants!) and should be calibrated with this in mind.

Disclaimer

MigClim is free for non-commercial research and educational purposes and comes with absolutely no warranty, expressed or implied. MigClim, this user guide as well as the accompanying test data all come "as is" and without any warranty of support from its author.

This being said, I'm happy to hear about your experience, bug reports, critics, etc... You can contact me by e-mail at: Robin.Engler [-at-] gmail.com.

Part 2: The MigClim model explained

The basic principle of the MigClim model

MigClim is a cellular automaton implemented within the ArcGIS® software (ESRI Inc., Redlands, USA). The cells (hereafter pixels) are square and record various values such as pixel occupancy status, habitat suitability, reproductive potential or when it was last colonised. To simulate dispersal under environmental change (e.g. climate change), MigClim requires the following inputs: a map defining the species' initial distribution, maps picturing landscape fragmentation (i.e. barriers to dispersal and permanent unfavourable locations), the species' dispersal parameters and a series of maps indicating how the distribution of potentially suitable habitats evolves as climate changes.

A first important point to understand about the MigClim model is that it does not generate habitat suitability maps itself. Thus, before being able to use MigClim, you have to generate your habitat suitability maps using another software, for instance the BIOMOD package (see Thuiller *et al.* 2009 – *Ecography* and http://r-forge.r-project.org/projects/biomod) that works with the free R software (www.r-project.org).

The basic principle of the MigClim model is the following: The user gives an initial species distribution (the starting point of the simulation) and one or more habitat suitability maps (maps indicating which pixels are suitable for the species and which are not, at one or several points in time). Using this data and the information about the dispersal ability of the species, MigClim will simulate the dispersal of your species and produce a potential distribution map that accounts for dispersal limitation.

One important concepts to understand about the MigClim model is what the N_{Disp} and N_{HSmap} parameters are and how they work (keep reading and these parameters will be explained, see also the model flow chart on page 15). When modeling the dispersal of species, the concept of time has necessarily to be considered, i.e. dispersal generally happens in a given time. For instance, a species might disperse seeds once a year and it might take 5 years before this species starts reproducing. So how is time expressed in the MigClim model?

The basic time unit in MigClim is a ***dispersal event*** (also called a "dispersal step"). A dispersal event corresponds to one loop in model where dispersal is simulated (see the model's flow-chart below). During a dispersal event, each pixel that is occupied and able to reproduce is given the opportunity to colonize new, empty and suitable, pixels. Each occupied pixel will also increase its "age" by 1 time unit. In practice, a dispersal event will often be equal to one year.

Another concept that is important in respect to how MigClim deals with "time" is the concept of ***environmental change event*** (also called "environmental change step") . To understand it, one should remember that MigClim was primarily developed to refine projections of future potential distribution of plant species under climate change scenarios. An ***environmental change event*** thus corresponds to a change/update in the habitat suitability map that is used in MigClim to define which pixels can be colonized and which cannot. Such change/update in habitat suitability reflects the change in habitat suitability as a consequence of the environmental change (e.g. a climate change scenario).

Now, the way that ***dispersal event*** and ***environmental change event*** relate is that the ***dispersal event*** loop is nested within the ***environmental change event*** loop (see the model's flow-chart below). The number of environmental change loops to be performed in a simulation is N_{HSmap}, and the number of dispersal event loops to be performed within each environmental change loop is N_{Disp}. The total number of dispersal events that will be performed within a simulation is thus equal to $N_{HSmap} \times N_{disp}$.

In practice, it is generally convenient to set the number of ***environmental change events*** and the number of ***dispersal events*** so that each dispersal event corresponds to one year. Here is a more concrete example to show what I mean:

Let's assume that we want to model the potential distribution of a specie under a climate change scenario for the year 2100 and that we know its initial distribution at the year 2000. How do we proceed ?

A first option could be to simply model its potential distribution by 2100 and then simulate 100 dispersal events, one for each year from 2001 to 2100. In this case, we would set $N_{HSmap} = 1$ because we have only one habitat suitability map (i.e. the one giving the habitat suitability by the year 2100), and $N_{Disp} = 100$, because we want to run the simulation for 100 years. So, if we multiply: $N_{HSmap} \times N_{Disp} = 1 \times 100 = 100$ years, which is what we wanted to do (i.e. we wanted to run the simulation for 100 years, from 2001 to 2100).

While this option seems easy to implement, it is by no means an ideal one. The problem here is that we modeled the habitat suitability only for the year 2100, with no intermediate step. Thus, it is very possible that, by doing so, we create important gaps between the initial distribution of the species and its potential distribution by 2100. In fact, the gaps that we created might well be beyond the species' dispersal abilities and the model will thus tell us that the species will go extinct because it is unable to colonize any of the newly suitable habitat.

A better way to proceed is therefore to model the change in habitat suitability for smaller time intervals. For instance, we could decide to model the change in habitat suitability every 5 years. This means that, from 2001-2100, we have to update the habitat suitability 20 times (100/5 = 20). The number of ***environmental change events*** must hence be set to $N_{HSmap} = 20$. Since we choose to update the habitat suitability every 5 years, this means that we must set $N_{Disp} = 5$, so that each ***dispersal event*** corresponds to one year. If we multiply: $N_{HSmap} \times N_{Disp} = 20 \times 5 = 100$ years, which, again, is what we wanted to achieve. This time however, our simulation will likely be more correct since we implemented climate change as a series of 20 events of small magnitude each, rather than just one event of big magnitude as in the first method discussed above.

In the extreme case, we could also choose to set $N_{HSmap} = 100$ and $N_{Disp} = 1$. This means that we would update habitat suitability every year, and thus run only one dispersal event per habitat suitability map update. This of course would require to produce 100 habitat suitability maps.

The MIGCLIM parameters

In order to simulate the dispersal of a species, the MIGCLIM model has a number of compulsory and optional parameters. All of these parameters can be modified so that their value best suits the species one wishes to model. Here is a list of the different MIGCLIM parameters and their description.

***** denotes optional parameters.

Parameter name	Description
Initial Distribution	Map giving the initial distribution of the species to be modeled.
Habitat Suitability Map(s)	One or several maps indicating the habitat suitability of each pixel for the species at a given time. There must be as many habitat suitability maps as there are environmental change steps (N_{HSmap}).
Number of environmental change steps (N_{HSmap})	Number of environmental change events to be implemented in the simulation. Each environmental change event/step corresponds to an update in the habitat suitability map, thereby reflecting the change in habitat suitability due to the environmental change (e.g. climate change scenario). To run a simulation without environmental change, set this parameter to "0".
Number of dispersal steps (N_{Disp})	Number of times that a dispersal event is simulated during each environmental change event/step. In general, the number of dispersal steps should be chosen so that each dispersal event corresponds to one year (assuming the the species will disperse once a year, or can be modeled as such).
Dispersal kernel (P_{Disp})	Probability of a pixel to act as "source pixel" (i.e. pixel that can be a seed source, leading to the colonization of another pixel), given the distance between that source pixel and a "sink pixel" (i.e. the pixel to be colonized).
Reproductive potential kernel (P_{mat}) *	Probability of a pixel to act as a "source pixel" given the time since it was colonized.
Filter *	Map indicating which pixels are permanently unsuitable for the species. For instance, urban areas and lakes could be used as filters.
Barrier *	Maps indicating which pixels represent barriers to dispersal for the species. Barriers to dispersal are pixels that are not only unsuitable for the species to inhabit, but also impede seed dispersal through them. For instance, forested areas could be an example of barriers for some species.
LDD minimum distance *	Minimum distance that can be reached by LDD (Long Distance Dispersal) events.
LDD maximum distance *	Maximum distance that can be reached by LDD (Long Distance Dispersal) events.
LDD frequency (P_{LDD}) *	Frequency with which LDD (Long Distance Dispersal) are generated.
Extinction frequency (P_{Ext}) *	Frequency of random (stochastic) pixel extinction.
Vegetative Resilience	Time during which a pixel (i.e. a population) can remain in a "vegetative resilient" state.
Seed Bank Resilience ($P_{SeedBank}$)	Time during which a pixel (i.e. a population) can remain in a "seed bank resilient" state, and probability associated to the recovery of the pixel from that seed bank as a function of time ($P_{SeedBank}$ = Seed bank resilience kernel).
River dispersal *	This regroups a number of parameters that allow to model specific dispersal of a species along river features, in addition to its normal dispersal. The parameters of this options are the following: - Multiplication factor for river dispersal: This is the multiplication factor by which the "standard" short distance dispersal kernel is multiplied for river dispersal.

- Minimum LDD distance along river features.
- Maximum LDD distance along river features.
- LDD frequency along river features.

Road dispersal * This regroups a number of parameters that allow to model specific dispersal of a species along road features, in addition to its normal dispersal. The parameters of this options are the same as for river dispersal (but they can have different values). The only difference between river and road dispersal is that in river dispersal, the dispersal direction is only in one direction (i.e. downstream), whereas in road dispersal the dispersion can occur in both directions.

Slope dispersal * This parameters allows to modify the dispersal distance so that it becomes spatially anisotropic: the dispersal distance will be increased in downhill directions, and decreased in uphill directions. Using this option requires a Digital Elevation Model (DEM) of the study area. The factor of dispersal distance increase/decrease with distance can be customized as a function of slope.

Wind dispersal * This parameters allows to modify the dispersal distance so that it becomes spatially anisotropic: the dispersal distance will be increased in "with-wind" directions, and decreased in "against-wind" directions. Using this option requires a map that contains the wind direction for each pixel of the study area and another one that contains the wind strength for each pixel. The factor of dispersal distance increase/decrease can be customized as a function of wind speed.

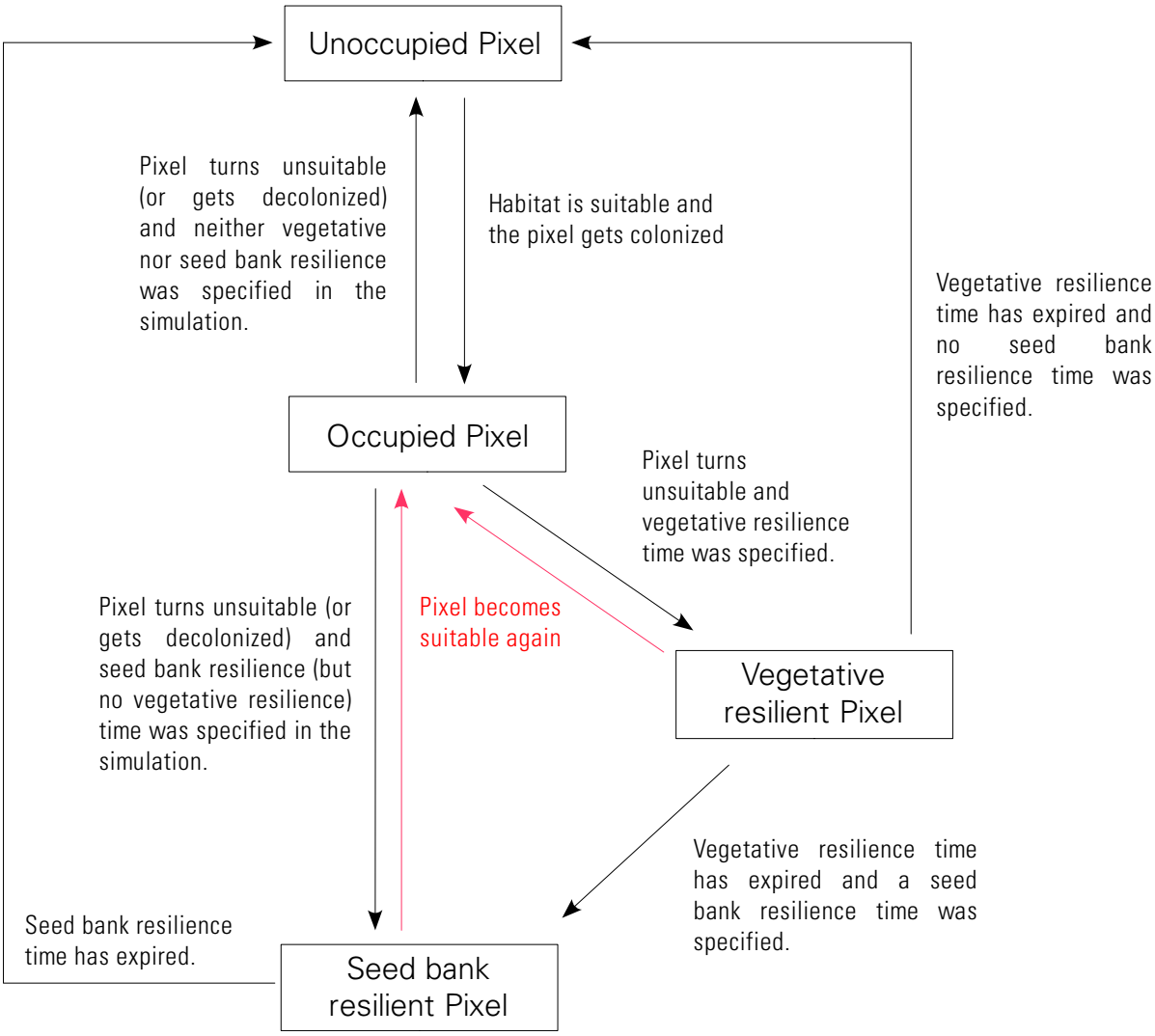
Pixel Size Factor A multiplication factor for the dispersal kernel (P_{Disp}) that can be used to compensate for difference in pixel sizes among different simulations. More information about this parameter can be found at the end of this user guide. Unless you understand what is does, I recommend leaving the value of this parameter to its default value of "1".

Important: The "wind dispersal" and "slope dispersal" options are still under development and might thus still contain some errors and bugs. (actually I haven't worked on these options since a while and I'm not even sure they are working in the newer versions of MigClim).

Pixel states in MigClim

In MigClim, pixels can be in four different states: ***unoccupied***, ***occupied***, ***vegetative resilient*** and ***seed bank resilient***. The description of each of these states is given here-below and the flow-chart illustrates the potential transitions between pixel states.

Note that the transitions shown below are only potential transitions: they do not necessarily occur as they might be subject to some supplementary conditions not mentioned in this flow-chart. For instance, transition from unoccupied to occupied occurs only if a suitable pixel actually gets colonized (which is a probabilistic process); transition from seed-bank resilient to occupied is also a probabilistic process and does not automatically occur (unless you set the probability to 100%, obviously).



Notes:

- *Vegetative and seed bank resilient stages are optional. They will only be used if you ask to do so.*
- *At the beginning of a MigClim simulation, pixels can only be in two states: unoccupied or occupied. Furthermore, occupied pixels will all be given full reproductive potential (P_{Mat} = 100%).*
- *Vegetative resilient pixels can only exist in simulations implementing environmental change (e.g. climate change). Seed bank resilient pixels can only exist in simulations implementing environmental change and/or random pixel extinctions ($P_{ext} > 0$).*

Unoccupied pixel: a pixel that is neither occupied by a population nor by a seed bank. In other words, it is completely empty.

Occupied pixel: a pixel that has been colonized by a population. The time since a pixel became occupied is continuously recorded and this information is used to determine how "mature" the pixel is through the user calibrated parameter P_{Mat}. Thus, depending on the kind of values that you defined for P_{Mat} , not all occupied pixels are equal: those that have been colonized since a longer time can have a higher reproductive potential.

Vegetative resilient pixel: This is a particular state that can only occur in simulations implementing environmental change (e.g. climate change). The idea behind vegetative resilience is that animals and plants can show some resilience to changes in their environment. When the conditions become suboptimal, they may still be able to survive in a given location but in a "resilient" state, where, for instance for plants, they are not able to reproduce anymore but are still surviving.

This is how vegetative resilience works: Imagine that a given pixel, which is occupied by a population during environmental change step *X*, becomes unsuitable at environmental change step *X+1.* If the vegetative resilience parameter is set to 0 (the default value), then the population will disappear from the pixel at the end of environmental change step *X+1.* Now, if a vegetative resilience time was defined, then the pixel will enter a "vegetative resilient" state.

Vegetative resilient pixels remain occupied by their population but with two important differences as compared to normal occupied pixels: first, pixels that are in a vegetative resilient state can no longer reproduce (i.e. they can no longer act as "source" cells to colonize other pixels). Second, the population in this pixel will no longer grow, which means that it will not gain any age anymore.

If, while a pixel is in vegetative resilience state, its habitat turns suitable again, then this pixel will switch back to an "occupied" state, where it is able to grow and reproduce (provided its $P_{Mat} > 0$). On the other hand, if a pixel has exhausted its vegetative resilience time without recovering, it switches to either "seed bank resilient" or "unoccupied" status, depending on whether the user has defined a *seed bank resilience time* or not. The time (measured, as always in MigClim, in *dispersal events*) during which a pixel can remain in vegetative resilience can be customized by the user to suit its particular species. Importantly, a pixel preserves its "age" while in vegetative resilience, and, when the conditions turn favorable again, it thus already has the "age" that it had before entering the vegetative resilience state (i.e. the age is not reset to zero).

Important: "vegetative resilience" is an optional parameter. Pixels will only be able to enter the a vegetative resilient state if you decided to specify a vegetative resilience time.

Seed bank resilient pixel: The idea of this pixel state is to simulate the fact that a plant population can produce seeds that will create a seed-bank in the ground. A seed bank will allow a population to recover without the assistance of other pixels, i.e. the pixel doesn't require a colonization event to become occupied, it can "auto-recover" by itself from its seed bank.

This state is entered by pixels that have exhausted their vegetative resilience time, or, if no vegetative resilience was specified, have become unsuitable due to environmental change or have become decolonized trough a random decolonization event.

If a pixel that possesses a seed bank becomes suitable again then it can "auto-recover" with a probability $P_{SeedBank}$, which depends on how long it has been in the seed bank resilient state. Once a pixel has exhausted its seed bank resilience time without recovering, it switches to "unoccupied" status. The time (measured in *dispersal events*) during which a pixel can remain in seed bank resilience can be customized. Generally, you will want to calibrate this parameter so that the value of $P_{SeedBank}$ (i.e. the probability that the pixel auto-recovers from its seed bank) decreases over time, so that the older a seed

bank becomes, the less likely it can recover a pixel.

Important: "seed bank resilience" is an optional parameter. Pixels will only be able to enter the seed bank resilient state if you decided to specify a seed bank resilience time.

Important: If a pixel has not reached full maturity at the time it turns "seed bank resilient", then it will only get a seed bank with probability P_{Mat} (which depends on its age = time since colonized). If the pixel has failed to build a seed bank, it will directly return to the "unoccupied pixel" status.

Important: The difference between "seed bank resilience" and "vegetative resilience" are mainly twofold.
First, the vegetative resilient pixels do keep their population, while the seed bank resilient pixels do not. This means that, if the pixel turns suitable again, a pixel that was in vegetative resilience will simply keep going as if nothing ever happened (e.g., if that pixel was 10 years old and already had a fully mature population, then it will immediately recover as a pixel that is 10 years old and fully mature). If a pixel that was in seed bank resilience state recovers, it starts again from the age "0", as if it just had been colonized (i.e. even if that pixel was 10 years old and fully mature, the population is completely lost).
Second, the vegetative resilient pixels do not have a recovery probability like seed bank resilient pixels have (i.e. $P_{SeedBank}$). Thus, while a vegetative resilient pixel will always recover if its habitat becomes suitable again, a seed bank resilient pixel will only recover with a probability $P_{seedBank}$. The probability $P_{seedBank}$ can be customized by the user and can have a value that changes over time (i.e. usually one will want to make the value of $P_{seedBank}$ decrease as the seed bank becomes older, reflecting that an older seed bank is less likely to lead to the recovery of a pixel).

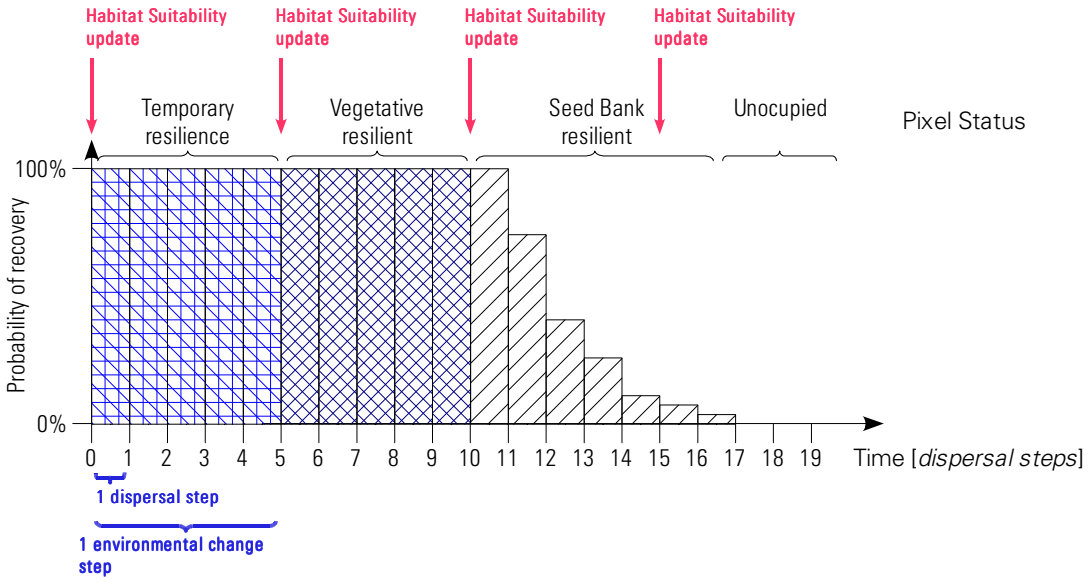
Important: Seed-bank resilience always starts after vegetative resilience, so if you indicate for instance 5 years of vegetative resilience and 5 years of seed-bank resilience, the 5 years of seed-bank resilience would start at the end of the 5 years of vegetative resilience. To give another example, in the figure just above this paragraph, the settings are: vegetative resilience = 5 years, seed-bank resilience = 7 years.

Temporarily resilient pixels: This pixel state has not been mentioned so far for the reason that it is hardwired into MIGCLIM and thus you cannot modify it. This is simply an temporary state of a pixel when it becomes decolonized due to environmental change (and thus this state only exists in simulations implementing environmental change). In this state, pixels continue to behave just as if they were fully colonized: they can grow and reproduce if they are mature enough. This state only lasts until the end of the environmental change step during which the pixel became decolonized due to environmental change. The idea of this state is that the transition from suitable habitat to unsuitable is not a discrete process, but a more continuous one. As such, the temporarily resilience state offers the opportunity to a pixel to still disperse during the environmental change step in which it becomes unsuitable.

At the end of each environmental change step (i.e. before the habitat suitability is updated), all temporarily resilient pixels are suppressed and turn into unoccupied pixels, or vegetative resilient or seed-bank resilient if the used has chosen to implement these options.

An illustration of the transitions between pixel states

This description of the transition between pixels states (see previous section for explanations on each pixel state) is mainly relevant to simulations that are implementing environmental change (e.g. climate change). Note that the values for environmental change step and vegetative resilience that I use in this example are the same than those that we will use in our test simulation later on ("Part 3: Hands-on Example with a MIGCLIM Test Simulation"). Seed-bank resilience is not exactly the same but almost (7 years instead of 5).



At time zero in the example above, a colonized pixel turns unsuitable (due to environmental change, indicated by the red arrow). As a result, the pixel turns into a ***Temporarily Resilient*** state. In this state, it continues to behave as if it was normally colonized until the end of the environmental change step (blue section of the graph; until dispersal step 5). Therefore you don't really need to pay a lot of attention to this particular pixel state, and the most important to understand is that a pixel remains able to colonized new pixels until the end of the environmental change step during which its habitat has turned unsuitable.

Important: Temporarily resilient state is hardwired into the model and will always last exactly until the end of the environmental change step during which a pixel turned unsuitable.

At the end of the environmental change step during which its habitat turned unsuitable, a pixel will leave the ***Temporarily Resilient*** state and change into either ***Unoccupied*** (if no vegetative resilience nor seed-bank resilience is implemented in the simulation), ***Vegetative Resilient*** or ***Seed-Bank Resilient*** (if no vegetative resilience was implemented).

Let's assume here that we have chosen to implement environmental change (i.e. update in habitat suitability) every 5 years (= 5 dispersal steps), that we gave our species a vegetative resilience time of 5 years and a seed bank resilience of 7 years. Thus, after our pixel has left its temporarily resilient state, it will become a ***Vegetative Resilient*** pixel. In this state, the pixel cannot colonize other pixels and does not gain age anymore but it remains "alive" so to say. If during the time (5 years in our example) in which it is in vegetative resilience the habitat turns favorable again, the pixel will recover to an ***Occupied*** state with 100% probability and will furthermore have kept its "age" (e.g. if the pixel was say 20 years old and fully mature, it will keep going from there on). Note that, because we have implemented environmental change every 5 years and a vegetative resilience of 5 years, our pixel has only 1 opportunity to recover from vegetative resilient to occupied: in the first year when the habitat suitability map is updated. In the other 4 years, dispersal is simulated but habitat suitability remains unchanged so if the pixel did not recover in the first year, it will not be able to do so during the other 4 neither.

If the habitat has not turned favorable again, then the pixel will change into ***Seed-Bank Resilient*** state. In fact, it is a little more complicated than that because, if the pixel has not reached full maturity, the seed bank is only built with a probability equal to its value of P_{Mat} which depends on its age (if the pixel has not even reached its initial maturity age then no seed bank is available and it turns into an unoccupied pixel). If a seed-bank was successfully produced, then the pixel remains in its seed-bank resilient state for up to 7 years (the time we have defined here for our imaginary simulation). Thus, during

the next 7 years (= dispersal events/steps), if its habitat is suitable, the pixel will attempt to recover to an occupied state. But, unlike when it is in vegetative resilience, the probability of recovery is now user defined and depends on the resilience kernel that is given as input by the user. Thus, unlike in vegetative resilience, the probability of recovery is not necessarily 100% anymore. If the pixel is able to recover during its seed-bank resilience time, it will turn into an ***Occupied*** pixel but with an age of "0" (i.e. it needs to start again its growth from scratch). If its habitat did not become suitable again or if the pixel was unable to recover from its seed-bank, it turns into an ***Unoccupied pixel***.

Dispersal simulation flow in in MIGCLIM

Dispersal is simulated through a number of decisions that are taken, for each pixel, during each dispersal event. The numbers correspond to those found on the next page's flowchart figure:

1. Does the target pixel represent a suitable habitat? Is it unoccupied?
2. If point 1 is answered positively, the number *n* of source pixels within the specified dispersal distance is computed. Source pixels are already occupied pixels that can act as seed sources to colonise a target pixel. Optionally, a barrier layer can be given to prevent dispersal through those pixels being part of the barrier. If a barrier pixel is found between the target and a source pixel, the source pixel is ignored. Barriers can be used e.g. to prevent a strictly grassland species to disperse through forests.
3. If $n > 0$, the target pixel becomes colonised with the combined probability P_{Col} (Eq. 1):

$$P_{Col} = 1 - \prod_{i=1}^{n} (1 - P_{Disp_i} \times P_{Mat_i}) \qquad \text{(Eq. 1)}$$

   Where $P_{Disp\ i}$ is a probability function of the distance between the target pixel and source pixel *i* and reflects the fact that colonisation probability decreases over distance. $P_{Mat\ i}$ is a probability that is function of the time since the source pixel *i* became occupied and represents the increase in reproductive potential of source pixel *i* over time. P_{Mat} can be used to represent time for individuals to reach reproductive maturity and, more globally, the increase of a population's reproductive potential due to an increase in the number of individual plants within a pixel over time. P_{Disp} and P_{Mat} are implemented as discrete functions and can easily be modified to fit any shape of seed dispersal curve and increase of reproductive potential over time.
4. Optionally, long distance dispersal (LDD) and stochastic extinction events can be added to the simulation. LDD events are generated from source pixels with a probability $P_{LDD} \times P_{Mat}$ in a random direction and at a random distance within a user-defined range. If the pixel reached by the long dispersing seed is potentially suitable (satisfying point 1), it becomes colonised. LDD events are not affected by barriers. Stochastic extinctions with probability P_{Ext} can also be defined to simulate random extinction of colonised pixels.
5. Steps 1 to 4 are repeated a number of times (N_{Disp}), typically set so that each repetition corresponds to one year.
6. Pixels that are no longer suitable due to changes in environmental conditions have their values reset to zero. Pixels that become unsuitable are reset only after the dispersion stage occurred (steps 1 to 5), because it is assumed that the change of a habitat from suitable to unsuitable is not a discrete but a continuous process. Thus, organisms inhabiting a pixel still have the potential to disperse during the step when the pixel turns unsuitable.
7. Steps 1 to 6 are repeated N_{HSmap} times. In each repetition, the habitat suitability is updated to

reflect environmental change (e.g. climate change). Simulations without environmental change can be performed by setting $N_{HSmap} = 1$.

Note: This description included only the most basic parameter available in MigClim. Additional parameters available include vegetative and seed bank resilience time, post-dispersal survival and/or habitat invasibility, anisotropic dispersal to simulate dominant winds or slope, as well as specific dispersal along certain features, such as roads and rivers. Some of these options are presented in the test simulation (see "Part 3 – Hands-on example with a MigClim test simulation" of this user guide) and some are presented in the section that describes additional MigClim parameters and options (see "Part 4 – Additional options available in MigClim" of this user guide).

Important: Because the smallest modelling unit in the model is a population in a pixel, parameter values must represent values for a whole population, not for a single individual. All parameters can be modified to best fit the known dispersal characteristics of the modelled species.

An illustration of the transitions between pixel states

This description of the transition between pixels states (see previous section for explanations on each pixel state) is mainly relevant to simulations that are implementing environmental change (e.g. climate change). Note that the values for environmental change step, vegetative resilience and seed-bank resilience that I use in this example are the same than those that we will use in our test simulation later on ("Part 3: Hands-on Example with a MigClim Test Simulation").

When a colonized pixel turns unsuitable (due to environmental change), it automatically turns into a ***Temporarily Resilient*** state. In this state, it continues to behave as if it was normally colonized until the end of the environmental change step. Therefore you don't really need to pay a lot of attention to this particular pixel state, and the most important to understand is that a pixel remains able to colonized new pixels until the end of the environmental change step during which its habitat turned unsuitable.

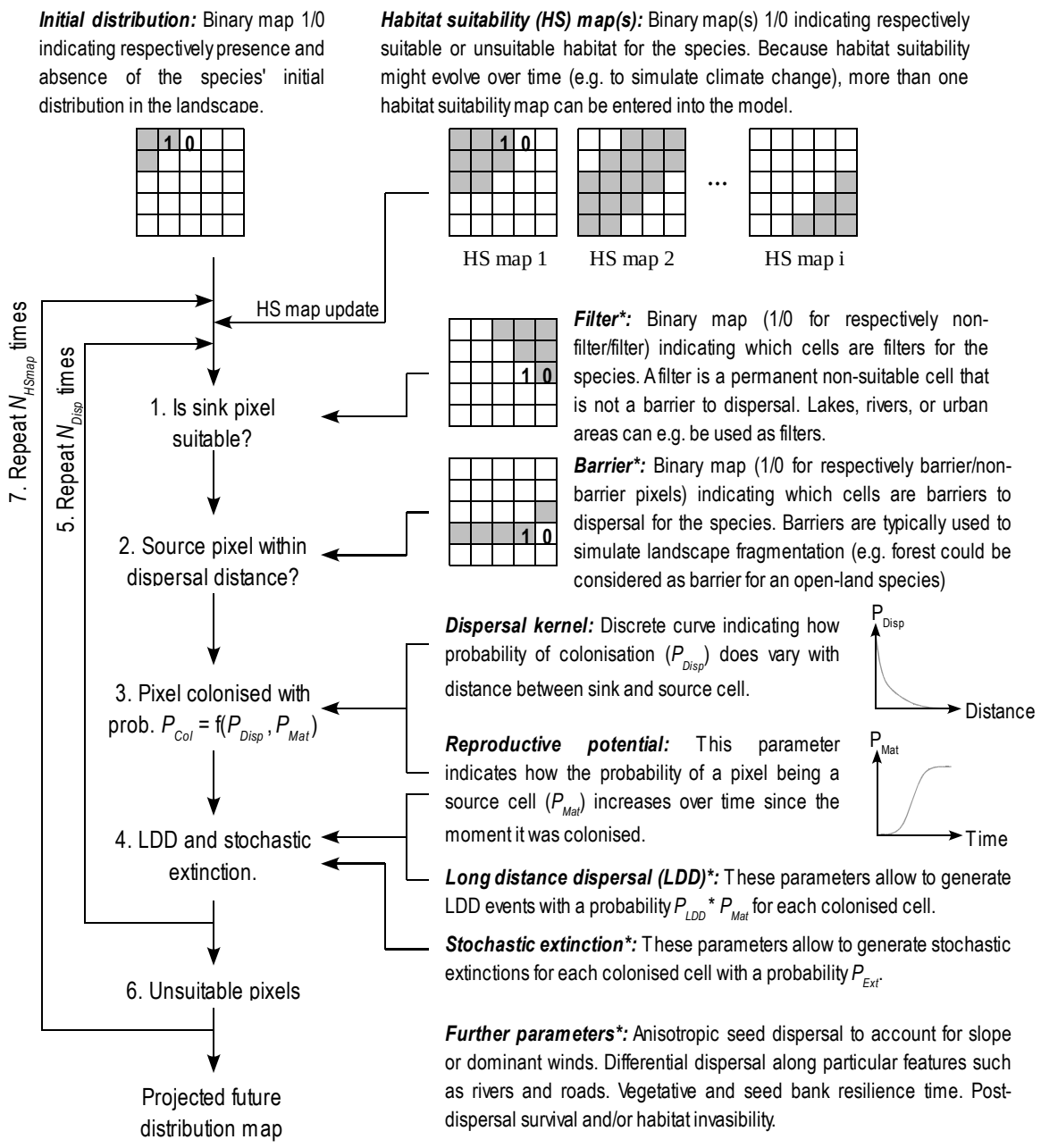
At the end of the environmental change step during which its habitat turned unsuitable, a pixel will leave the ***Temporarily Resilient*** state and change into either ***Unoccupied*** (if no vegetative resilience nor seed-bank resilience is implemented in the simulation), ***Vegetative Resilient*** or ***Seed-Bank Resilient*** (if no vegetative resilience was implemented). Note that the temporarily resilience state always lasts exactly until the end of the current environmental change step and that this cannot be modified by the user.

Let's assume here that we have chosen to implement environmental change (i.e. update in habitat suitability) every 5 years, that we gave our species a vegetative resilience time of 5 years and a seed bank resilience of 5 years as well. Thus, after our pixel has left its temporarily resilient state, it will become a ***Vegetative Resilient*** pixel. In this state, the pixel does not gain age anymore but it remains "alive" so to say. Thus, if during the 5 years in which it is in vegetative resilience the habitat turns favorable again, the pixel will recover to an ***Occupied*** state with 100% probability and will furthermore have kept its "age" (e.g. if the pixel was say 20 years old and fully mature, it will keep going from there on). Note that because we have implemented environmental change every 5 years and a vegetative resilience of 5 years, our pixel has only 1 opportunity to recover from vegetative resilient to occupied: in the first year when the habitat suitability map is updated. In the other 4 years, dispersal is simulated but habitat suitability remains unchanged so if the pixel did not recover in the first year, it will not be able to do so during the other 4 neither.

If the habitat has not turned favorable again, then the pixel will change into a ***Seed-Bank Resilient*** state. In fact, it is a little more complicated than that because, if the pixel has not reached full maturity, the seed bank is only built with a probability equal to its value of P_{Mat} which depends on its age (if the pixel has not even reached its initial maturity age then no seed bank is available and it turns into an unoccupied pixel). If a seed-bank was successfully produced, then the pixel remains in its seed-bank resilient state for up to 5 years (the time we have defined here for our imaginary simulation). Thus, during

the next 5 years (= dispersal events/steps), if its habitat is suitable, the pixel will attempt to recover to an occupied state. But, unlike when it is in vegetative resilience, the probability of recovery is now user defined and depends on the resilience kernel that is given as input by the user. Thus, unlike in vegetative resilience, the probability of recovery is not necessarily 100% anymore. If the pixel is able to recover during its seed-bank resilience time, it will turn into an ***Occupied*** pixel but with an age of "0" (i.e. it needs to start again its growth from scratch). It habitat did not become suitable again or if the pixel was unable to recover from its seed-bank, it turns into an ***Unoccupied pixel***.

Simplified flow chart of a MIGCLIM simulation. The numbers refer to those given in the text above



Part 3: Hands-on example with a MigClim test simulation

The aim of this section is to illustrate the use of MigClim by going through a step-by-step example that will show how to use MigClim with the provided test data. Along the way, the different MigClim parameters will be explained in details.

To run the simulation that is described in this tutorial, you will need to have the test data that came with this used guide. If you do not have the test data, it can be downloaded, along with this user guide and the MigClim model from http://www.unil.ch/ecospat/page47587_en.html (if this first link doesn't work, you can just go to www.unil.ch/ecopsat and look around, the download page shouldn't be further than a few clicks away).

If you haven't done it so far, I strongly encourage you to read the first part of this user guide. Knowing the basic principles and the various parameters of MigClim will help you better understand what is going on in this example.

Note: In MigClim, some parameters are compulsory while others are optional. In the following example we will use some of the optional parameters, but not all. There is another section, at the end of this user guide (see "Part 4: Additional Options available in MigClim"), that provides explanations on those optional parameters that are not covered in the example that will follow.

Getting started with the tutorial data

After opening the compressed ".zip" file that contains the tutorial data, copy all of it to a folder on your hard drive. For the purpose of this tutorial, I suggest to create the following folder "C:\MigClim\TestData" to host the tutorial data.

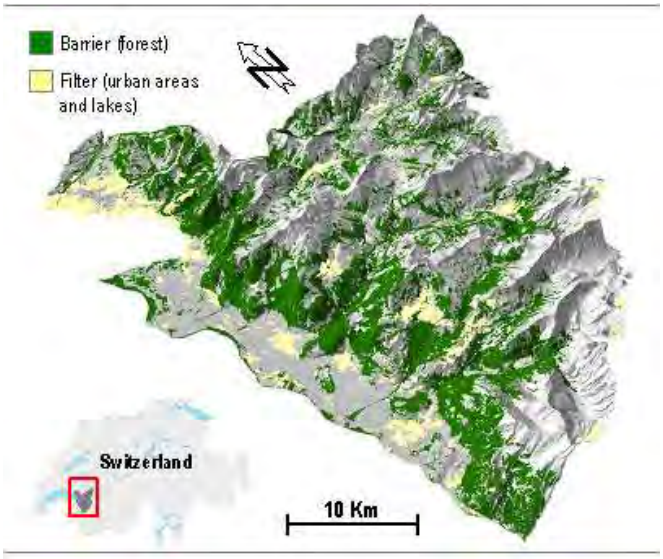
You are of course free to copy the data to any other location but make sure that the data path **does not contain any space!** E.g., "C\MigClim\Test Data\" or "C\Program Files\MigClim\TestData" are not an option as both these paths contain a space. Please do also avoid using all sort of special characters such as "$" or so in the path.

Got everything copied? Excellent, open the "MigClim.mxd" ArcMap project and let's get started, I'm sure you can hardly wait!

Test Data Description

Before things get too serious with our hands-on example (next section), let's start with a little overview of the provided test data. The aim of this section is to provide you with the general idea of the test simulation that we will run and describe the different files that compose the test data. Going through the description of the test data, I will also point out some important requirements of MIGCLIM's input data in general.

In our test example, which uses the test data that accompanies this user guide, our aim will be to model the dispersal of a species under a climate change scenario over a period of 50 years, from 2001 to 2050. The study area is a part of the (beautiful) Swiss Alps.



First, we will have a look at the different raster datasets (grids) that are part of the test data accompanying this user guide. Notice that you will have to use your ARCMAP® or ARCCATALOG® applications to view these grids. All raster datasets (grids) that are part of the test data have a pixel resolution of 50 meters and have exactly the same extent.

Important: It is a requirement for MIGCLIM to work correctly that all the grids (raster datasets) that you use for a given simulation have all exactly the same extent and pixel size.

Important: All inputs grids for MIGCLIM must be in the ESRI GRID format. They must all have exactly the same cell-size/pixel-size and exactly the same extent (i.e. the same number of rows and columns).

initial_dist: This is a raster map that represents the initial distribution of our test species.

- ➔ Pixel value of 1 = occupied pixel
- ➔ Pixel value of 0 = non-occupied pixel

As can be seen from the figure below, the initial distribution of our species is mainly restricted to the lowland area to the west of the study area.

initial_dist

Grey pixels have a value of 0, indicating they are unoccupied.

Orange pixels have a value of 1, indicating they are occupied by the species (i.e. initial distribution)

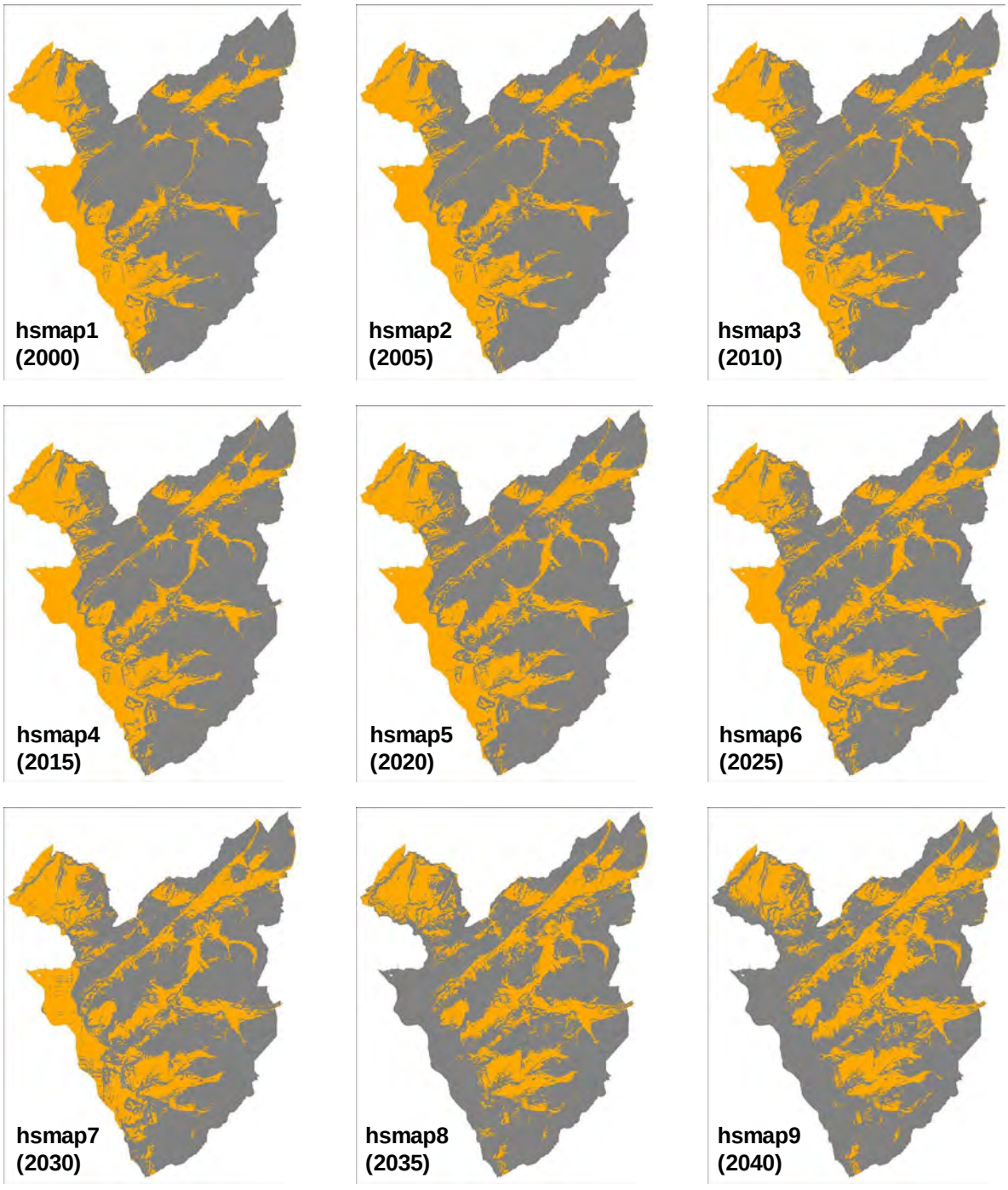
Important: pixel with a "0" value simply indicate that the species is absent from those pixels at the beginning of the simulation. However, it doesn't mean that those pixels cannot be suitable for the species at the beginning or later on in the simulation. This information has to be given by the habitat suitability maps.

Note: In this example data, study area is delimited by "NoData" values (this makes it look nice). However, I would advice not to use "NoData" values in any of your input grids but only values of "0" and "1".

hsmap1, hsmap2, hsmap3, ...hsmap10: These are raster maps indicating the habitat suitability of pixels at different periods in time. Just like the species' initial distribution map, these grids are binary and contain only two possible values: 1 = suitable pixels, 0 = unsuitable pixels.

In the particular case of the test data, "hsmap1" represent the habitat suitability for the year 2005, "hsmap2" the habitat suitability for the year 2010, "hsmap3" for 2015, … and so on until "hsmap10" that represents habitat suitability for the year 2050. The entire map series thus reflects the evolution of habitat suitability with climate change from 2005 to 2050 for our species.

Habitat suitability maps 1 through 9 are given in the figure below and show the evolution of the habitat suitability from 2005 ("hsmap1") to 2045 ("hsmap9"). "hsmap10" is not shown. Orange pixels have a value of 1 and indicate suitable habitat, which can potentially be colonized/occupied. Grey pixels have a value of 0, indicating unsuitable habitat that cannot be colonized/occupied.



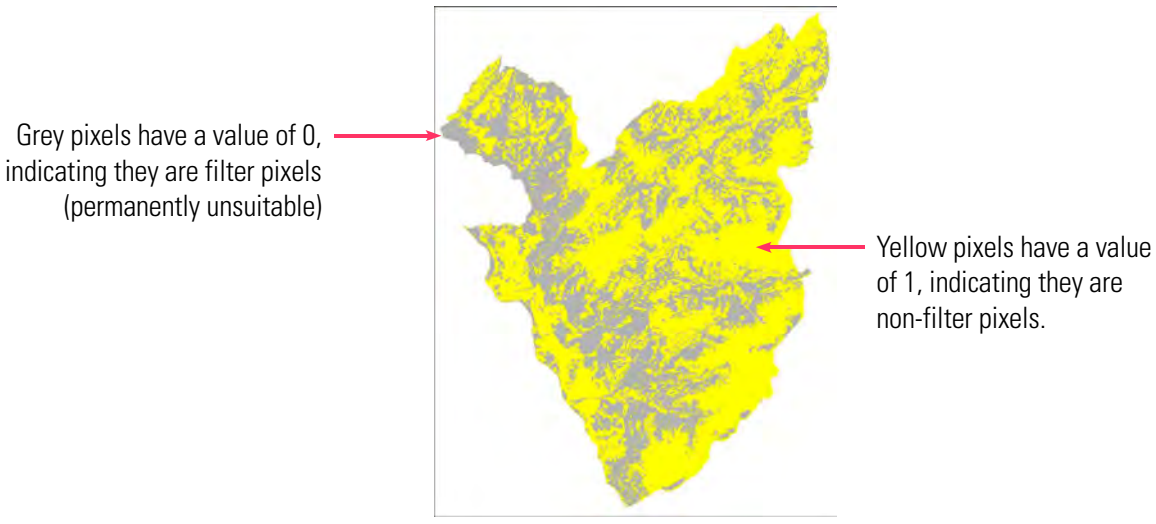
Important: *In this example, the habitat suitability is described using binary values: 1 indicating suitable habitats, 0 non-suitable habitats. However, MigClim can also deal with continuous values of habitat suitability (from 0-100). More details on this advanced option of MigClim are given later in this manual.*

test_barrier: a raster representing barriers to dispersal for the species (1 values = barrier pixel, 0 values = non-barrier pixel). In this case the barrier pixels correspond to forested areas as we will assume that our species cannot disperse through forests.

Note: barrier pixels remain barriers during the entire simulation and, at this stage, cannot be updated during the simulation. It could be argued that this is not fully realistic as barriers pixels might also evolve with time. For instance we could imagine scenarios where forests are displaced over time.

test_filter: a raster representing permanently unsuitable habitat (0 values = filter pixel, 1 values = non-filter pixel). In this case the filter pixels correspond to forested and urban areas as we will assume that our species cannot live in such environments.

Note: filter pixels remain unsuitable during the entire simulation and cannot be updated during the simulation. It could be argued that this is not fully realistic as filter pixels might also evolve with time. For instance, we could imagine scenarios where urban areas grow and forests move over time.

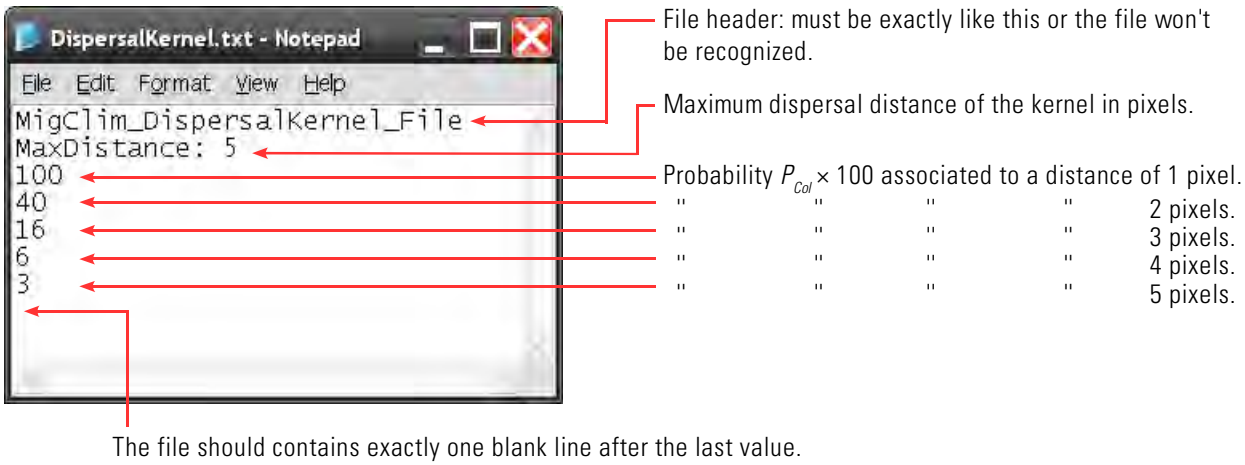


Important: What is the difference between the "barrier" and "filter" parameters? In both cases, pixels belonging to a barrier or a filter will remain permanently unsuitable for the species during the entire simulation. But filter pixels allow the species to disperse through them, whereas barrier pixels block dispersal (except long distance dispersal events). In other words, if your filter map and barrier map are the same, then you could use only the barrier map and the outcome would be the same.

Note: barrier pixels are identified by a value of "1" while filter pixels are identified by a value of "0". Make sure not to confuse them!

DispersalKernel.txt: a text file containing the probability values of P_{Col} associated to each distance. Probabilities must be given as values of P_{Col} × 100, and are thus on a scale going from 0 to 100. For instance, 0 means P_{Col} = 0, 100 means P_{Col} = 1 and, and 25.5 means P_{Col} = 0.255. Values with decimals are possible.

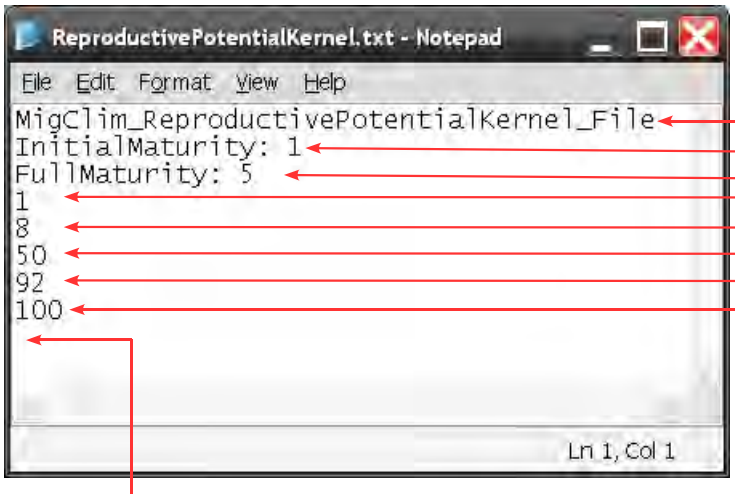
This file is constructed as follows:



Using this file, our species will be able to disperse over a distance of up to 5 pixels. The pixels located at a distance of 1 pixels (i.e. directly adjacent) to a source pixel will have a P_{Col} value of 100% (P_{Col} = 1), pixels located at a distance of 2 pixels a P_{Col} value of 40% (P_{Col} = 0.4), and so on until pixels located at a distance of 5 pixels that have a P_{Col} value of 3% (P_{Col} = 0.03).

ReproductivePotentialKernel.txt: a text file containing the probabilities of P_{Mat} associated to each pixel "age" from initial maturity to full maturity. Note that pixel age is measured in "dispersal events" units, but generally one dispersal event corresponds to one year.

Probabilities must be given as values of P_{Mat} × 100, and are thus on a scale going from 0 to 100. For instance, 0 means P_{Mat} = 0, 100 means P_{Mat} = 1 and, and 50 means P_{Mat} = 0.5. Values with decimals are possible.



File header: must be exactly like this or the file won't be recognized.

Time of initial maturity in dispersal steps.

Time of full maturity in dispersal steps.

Probability P_{Mat} × 100 associated to pixels of age 1.

" " " age 2.

" " " age 3.

" " " age 4.

Probability of P_{Mat} × 100 associated to pixels of age 5. This last value corresponds to full maturity and must thus always be equal to 100 (i.e. P_{mat} = 1).

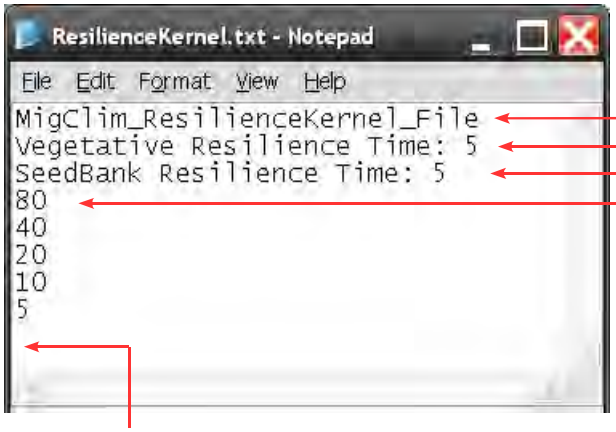
The file must contain exactly one blank line after the last value.

Using this file, a pixel (i.e. a population) of the modeled species will start being mature at age 1, which means 1 dispersal event after the pixel became colonized. Pixels will reach full maturity (100%; P_{Mat} = 1) after 5 dispersal events.

After one dispersal event, P_{Mat} will have a value of 0.01 (1%), after two dispersal events, P_{Mat} will have a value of 0.08 (8%), and so on until the age of five (5 dispersal events) where P_{Mat} will have reached 100%.

Important: the value of P_{Mat} associated with the last time (i.e. full maturity) must always be equal to 100%.

ResilienceKernel.txt: a text file containing the time during which a population (i.e. a pixel) can remain in vegetative resilience, the time of seed bank resilience, as well as the values of the probability P_{Recol} × 100 associated to each of these times. As always, the time is expressed in units of dispersal events (this is the time unit that is always used in MigClim).



File header: must be exactly like this or the file won't be recognized.

Vegetative resilience time in dispersal steps.

Seed bank resilience time in dispersal steps.

Probability P_{Recol} × 100 associated to a time of 1 dispersal event after a pixel became decolonized.

The file must contain exactly one blank line after the last value.

Using this file, our species has a vegetative resilience time of 5 years (i.e. 5 dispersal steps) and a seed bank resilience of 5 years. The seed bank initial recovery potential is of 80% but is halved with each year.

Remember that vegetative resilience only starts at the end of the environmental change step in which a pixel becomes unsuitable, which means in our case after 5 years (because during the 5 first years a pixel is in temporarily resilience; see also the graphic on page 12 of this manual)

Seed-bank resilience only starts at the end of vegetative resilience, so the last opportunity for recovery through seed-bank resilience is in fact after 5+5+5=15 years (5 years temporarily resilience + 5 years vegetative resilience + 5 years seed-bank resilience).

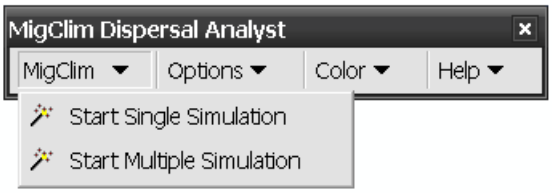
The MigClim toolbar

MigClim toolbar

Installing MigClim on your computer should give you access to the following toolbar within the ArcMap application. The different buttons of the toolbar are illustrated and discussed in this section.

MigClim menu

Two buttons are available in this menu: "Start Single Simulation" and "Start Multiple Simulation".



Start Single Simulation launches a new MigClim simulation. "Single simulations" allows entering data through a number of user-friendly forms that explain and guide you through the different available parameters. This is probably the way you want to run MigClim in your first simulations. This is also the way we will run our test simulation in this user guide.

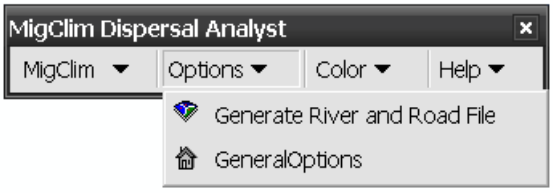
Start Multiple Simulations launches a new MigClim "Multiple Scenario" simulation. "Multiple Scenario" is a kind of batch mode that allow you to run an unlimited number of simulations automatically. Multiple simulations differ from single simulations in two aspects:

- The data input is not done through a series of forms but through a single text file, which is much faster once you have that input file ready.
- The text file can contain instructions for more than one simulations, making it easy to launch a whole series of simulations in just a few clicks.

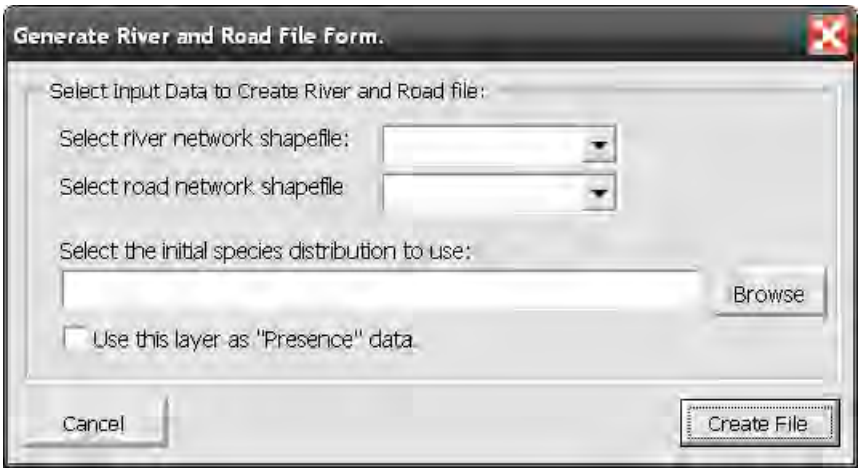
More information about the "Multiple Scenario" mode of MigClim are given at the end of this user guide in "Part 4: Additional options available in MigClim".

Options menu

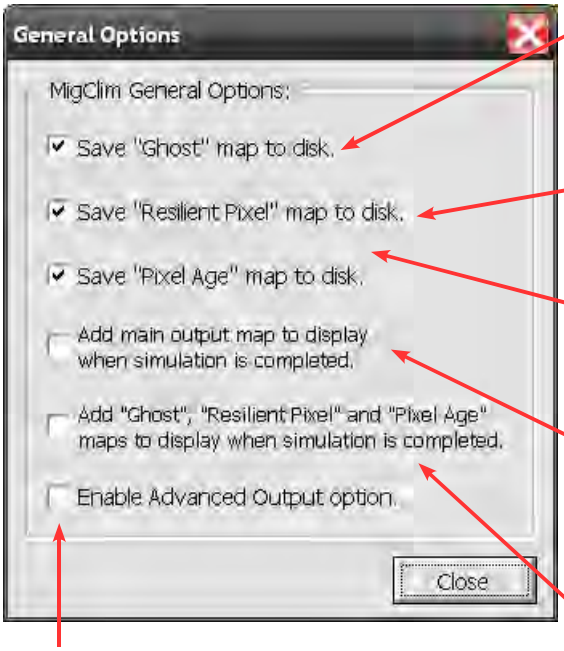
This menu contains two buttons: ***Generate River and Road File*** and ***General Options***.



The ***Generate River and Road file*** button will launch the following input form, that enables you to create a "River and Road" file. This type of file is necessary if you wish to include river and/or road dispersal into your simulations. More information on this option is given at the end of this user guide in "Part 4: Additional options available in MigClim".



The ***General Options*** button allows setting a number of options that affect the kind of output produced by MigClim simulations.



If checked, a "Ghost" map (raster) is produced as supplementary output. This raster allows you to see pixels that became decolonized during the simulation (see also the output interpretation section later in this tutorial for more details).

If checked, a "Resilient Pixel" map (raster) is produced as supplementary output. This raster allows you to see which pixels are in resilient state at the end of a simulation (see also the output interpretation section later in this tutorial for more details).

If checked, a "Pixel Age" map (raster) is produced as supplementary output. This raster allows you to see the "age" (time since colonized) of all pixels (see also the output interpretation section later in this tutorial for more details).

If checked, the main output raster is automatically added to your ArcMap project at the end of the simulation. I would advise not to enable this feature if you are running multiple simulations (because adding too many maps to the ArcMap view might generate problems).

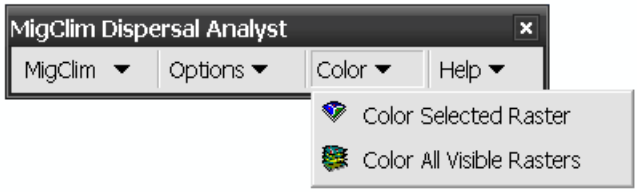
Check this box to get more detailed information in your output files. Not necessary for most users so leave it unchecked unless you know how to interpret the added results.

If checked, the "Ghost", "Resilient Pixel" and "Pixel Age" output raster are automatically added to your ArcMap project at the end of the simulation. I would advise not to enable this feature if you are running multiple simulations (because adding too many maps to the ArcMap view might generate problems).

Note: "Resilient Pixel" maps are only produced if the simulation implements environmental change, because if not, then the "Resilient Pixel" map is exactly the same than the main output.

Color menu

The "Color" buttons provide an easy way to set a nice color scale to MigClim output rasters. You can see an example of that later in this manual.



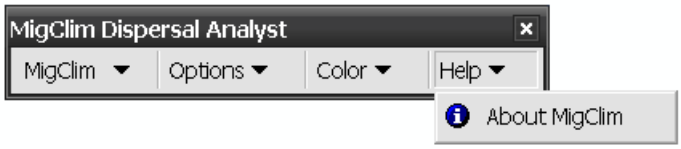
Color Selected Raster will color only the grid (i.e. raster layer) that is currently selected (highlighted) in the ArcMap "Table of Contents".

Color All Rasters will color all grids that are currently turned-on (visible) in the ArcMap "Table of Contents".

Note: the "Color" buttons will give you good results only on the rasters that are outputs of MigClim. *Make sure that you only have* MigClim *output rasters that are turned-on (i.e. are visible) before clicking on "Color All Rasters"*

Help menu

The ***About MigClim*** button provides information about how to cite MigClim in publications and how to report bugs, requests for improvements, your (many) complaints to its author (unlikely compliments are also kindly accepted :-).



MigClim Test simulation

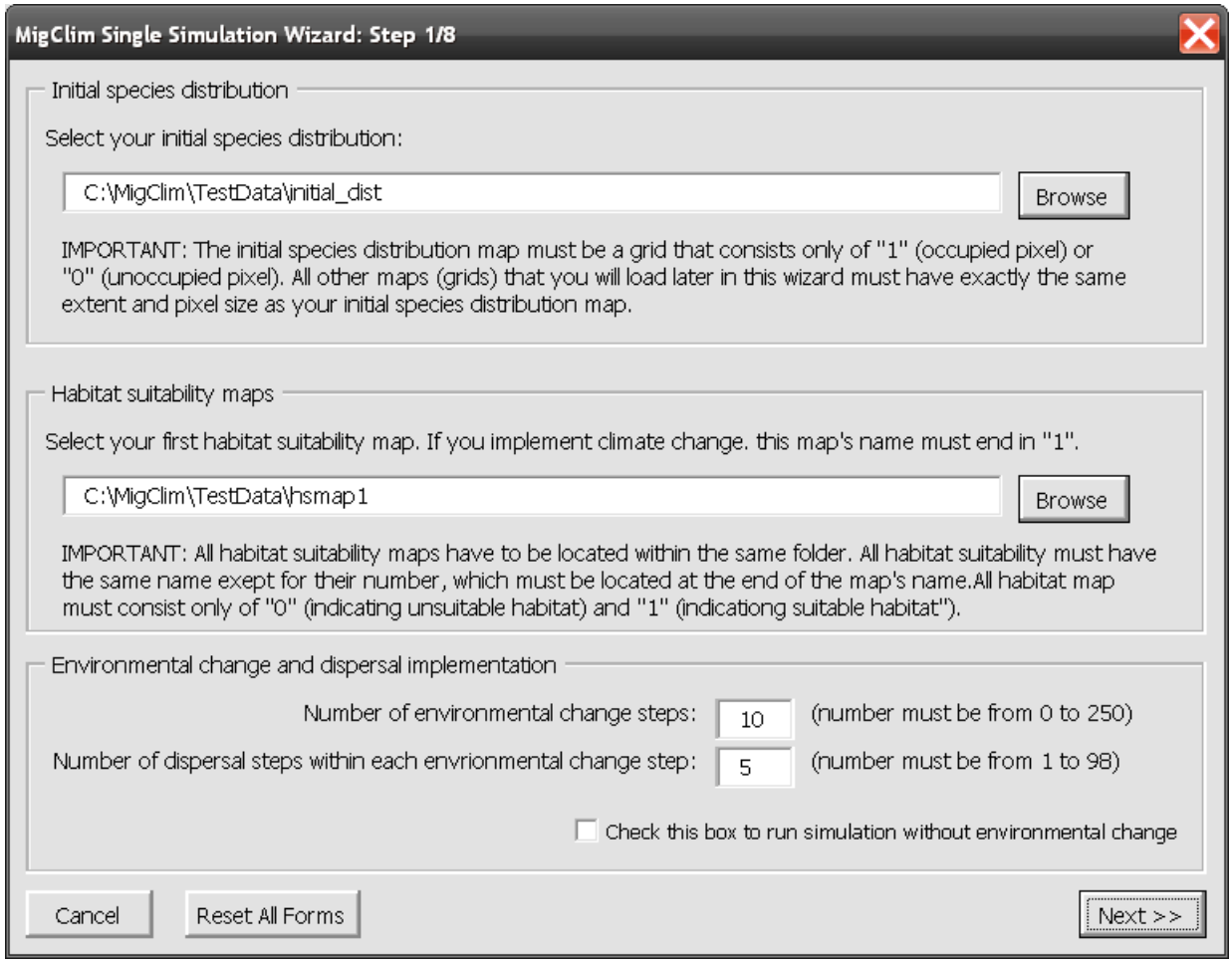
Now that we have explored the MigClim toolbar, let's start a simulation using the test data provided with this user guide.

Note: Don't get discouraged by the length of the tutorial for this test simulation. It is only so long because I added a lot of explanations to help you understand the parameters of MigClim. Once you will be familiar with the parameters, you will be able to use a "batch mode" that allows entering all parameter values via a text file in just a few clicks (see also "Part 4: Additional options available in MigClim" for mode details on the batch mode).

Note: Obviously, all the values that we will be using in this simulation are only examples and any resemblance with plants or animals that exist or have existed is purely coincidence ;-). In other words, you will have to work out your own parameters when using MigClim for your species.

Step 1: Launch the "Single Simulation Wizard" and specify species distribution parameters

In the MigClim toolbar, click the ***Start Single Simulation*** button. The first dialog box of the MigClim "Single Simulation Wizard" should show up as follows:



Step 1.1: Enter the your specie's initial distribution

First of all, you are asked to indicate your ***initial species distribution*** map. This map must be an ESRI GRID file that indicates which pixels are occupied by the species at the beginning of the simulation. The grid must contain only two kind of values: "0" indicating that the species is absent from the pixel, "1" indicating that the pixel is occupied by the species. The initial distribution map of our test data is unsurprisingly named "initial_dist", so click on "Browse" and locate this raster map on your computer. It should be in the "C:\MigClim\TestData" directory.

Note: "No Data" (instead or in addition of 0) values are supported but it is advised to avoid them nevertheless (unlike what I did in the test data!).

Step 1.2: Enter one or more habitat suitability map(s) for your species

Now you have to indicate the ***first habitat suitability*** map for your species. Let me elaborate a little on that: in order to simulate the dispersal of a species through the landscape, MigClim needs to know which pixels are favorable for the species to establish and which are not. The way that MigClim get this information from you is through "habitat suitability maps". These maps must be in ESRI grid format at contain values of "0" and "1" (0 = unsuitable habitat, 1 = suitable habitat).

Note: it is also possible to give finer values of habitat suitability, between 0 and 100, which allows to indicate a gradient of suitability, rather than just a binary value. See the "Part 4: Additional options available in MigClim" section at the end of this user guide for more information on this option.

Step 1.3: Implementing environmental change events and dispersal events

As we choose to implement environmental change over time (i.e. climate change in our case) in our simulation, you will agree that we need to have more than one habitat suitability map, since it is this succession of maps that will tell how the favorable habitat evolves over time. Because it would be tedious to have to enter the name of all of these maps manually, the trick is that all your habitat suitability maps **must have the same name, except for the number in which they end, and be located in the same folder**. It is the number at the end of each habitat suitability map's name that indicates the order in which they will be used and therefore indicating just the first of them is enough. This is why your **first habitat suitability map** (the one that you indicate in the form) **must end in "1"**.

In out test simulation, the habitat suitability maps are the following: hsmap1, "hsmap2", "hsmap3", "hsmap4", "hsmap5", "hsmap6", "hsmap7", "hsmap8", "hsmap9" and "hsmap10". They thus all have the same name "hsmap" and differ only by their final number. This naming convention **is very important and you have to respect it** in order to get your simulation to work.

If you plan to run a simulation without environmental change, then you don't need to worry about those rules because you will have only one habitat suitability map (habitat suitability is thus going to remain constant throughout the entire simulation).

Important: All habitat suitability maps must be located in the same folder than the first habitat suitability map. The first habitat suitability map has to end in "1" (e.g. "hsmap1"). The other habitat suitability maps must end with a number that indicates the order in which they will be used. The numbers of habitat suitability maps must be consecutive.

Important: All input grids, such as initial species distribution, habitat suitability maps or barriers to dispersal and filters (these will be discussed later in the example) ***must all have exactly the same extent, cell size and number of pixels****. In other words, all grids that you will ever use in one same MigClim simulations must be exactly superposable to each other. Should any of your grids not satisfy this criteria, your* ***results will be wrong****.*

Now that we indicated the locations of the initial species distribution raster and the habitat suitability rasters, we need to indicate the ***number of environmental change steps***. This parameter indicates to MigClim how many environmental change steps you wish to implement in your simulation. An "environmental change step" corresponds to a change (or "update") in habitat suitability map. For instance, in the test data that is provided, we have 10 habitat suitability maps. Each of these maps represent the habitat suitability of our species for a 5 years time period from 2005 to 2050 ("hsmap1" represents the habitat suitability for 2000-2005, "hsmap2" the habitat suitability for 2005-2010, and so on until "hsmap10" that represent the habitat suitability for the period 2045-2050). In other words, the 10 habitat suitability maps represent the evolution of habitat suitability of our species from 2000 to 2050 under a given climate change scenario.

In our tutorial example, the interval between two successive habitat suitability maps is of 5 years. You are of course free to use any other interval. For instance, you could implement a change in habitat suitability every year, every 10 years or every 20 years. It all comes down to what kind of data you have available and what kind of values you think does make sense: e.g. implementing climate change every year might not be ideal because annual variability cannot be easily predicted and hence it might be better to work with averages over a few years. On the other hand, implementing climate change only every 20 years might not be optimal either because it could result in changes that are too abrupt between two successive habitat suitability maps.

While you are free to choose any time interval between two environmental change events/steps, you will need to adapt the number of ***dispersal steps*** to the time interval you chose to have between ***environmental change steps***. A ***dispersal step*** basically corresponds to one dispersal event for your species. Usually this will corresponds to one year, because populations of most plant species will disperse their seeds once a year or can be modeled as such (this is probably also true for many animal species). What is important to understand here is that the chosen number of dispersal steps will repeat within each environmental change event (as explained earlier, see also the model's flowchart earlier in this document). For instance, in our tutorial simulation, we update the habitat suitability maps every 5 years, and therefore a environmental change event lasts 5 years. Thus, if we want to model the dispersal of our species on a yearly basis (i.e. one dispersal event per year), the number of dispersal steps should be set to "5", so that we have one dispersal event every year.

The total number of dispersal events that will occur during one simulation can be obtained by multiplying the number of environmental change steps by the number of dispersal steps. In our test data, we have the following values:

- Environmental change steps = 10
- Dispersal steps = 5

The total number of dispersal events simulated will thus be 10 × 5 = 50, which is well in line with what we wanted to achieve, namely modeling the species distribution over 50 years, from 2001 to 2050.

Note: How to run simulations without environmental change (e.g. climate change)?
Implementing environmental change in your simulation is not a requirement in MigClim. Indeed simulations without environmental change are even easier to implement because you don't need to have different habitat suitability maps but just one. An example of simulations without environmental change could be if you want to model the spread of an invasive species trough the landscape. The initial distribution map could be the point of introduction of the invasive species, and the habitat suitability map the potential suitable habitat for the species.
To run a simulation without environmental change, you have to to check the ***Run simulation without environmental change*** *box, which sets the number of environmental change steps to zero. The number of dispersal events is then simply the number of times you want to have a dispersal event for your species. For instance, if you want to simulate the dispersal of your invasive species over 50 years, and if your species disperses once a year, you would set the number of dispersal steps = 50.*

Note: Why does MigClim not generate habitat suitability maps itself and where can I get them from?
If you're reading this user guide, chances are that you already have some knowledge of species distribution modeling. Generating habitat suitability maps for present and future climatic conditions can be done in a large variety of ways that usually involve statistic modeling tools. Because of the large number of available technique and because there are already very efficient tools out there to generate such maps, MigClim does not do it. You thus have to generate the habitat suitability data yourself and convert them into ESRI GRID format before being able to use MigClim to simulate dispersal

Note: a future version of MigClim should be able to import the habitat suitability information directly from the output of the BIOMOD software (see Thuiller et al. 2009 – Ecography, and http://r-forge.r-project.org/projects/biomod/), or more generally from a simple text file, so that you won't have to convert all your data into ESRI GRIDs anymore!

Step 1.4: Moving on to the next form

That was it, we're done with the first form of the MigClim single simulation wizard and you can hit ***Next >>*** to move on. The ***Reset All Forms*** button resets all forms of the wizard (i.e. if you do another simulation in the same ArcMap session, the parameter values are kept in memory. If you wish to reset all parameters, hit that button). The ***Cancel*** button will logically exit the wizard.

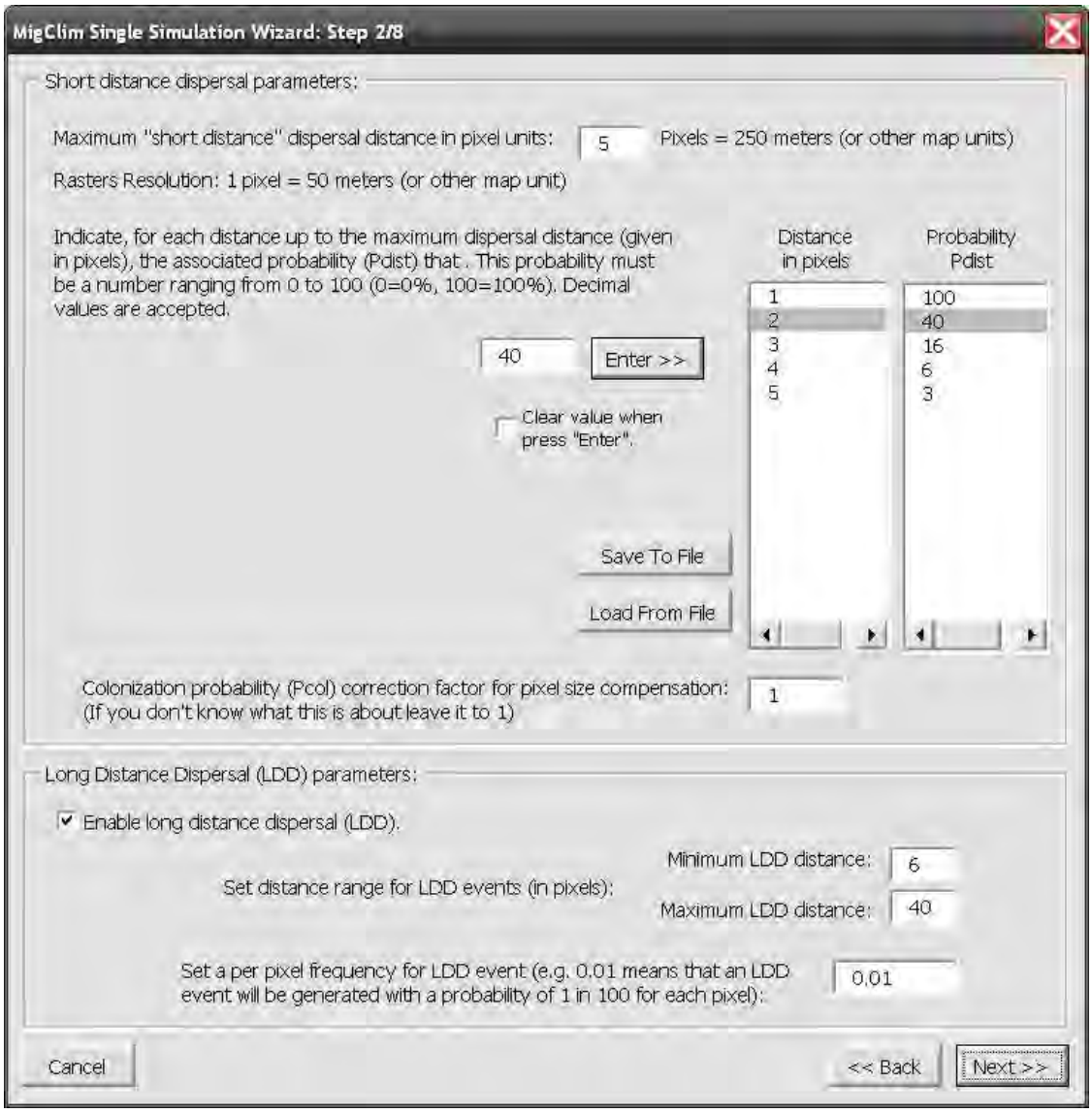
*Note: Once you hit **Next >>**, the information you entered will be verified. MigClim will notably verify that there are enough habitat suitability maps to accommodate for the number of environmental change steps that you have entered. This verification can sometimes take a little time (a few seconds or minutes) so please be patient.*

Step 2: dispersal parameters

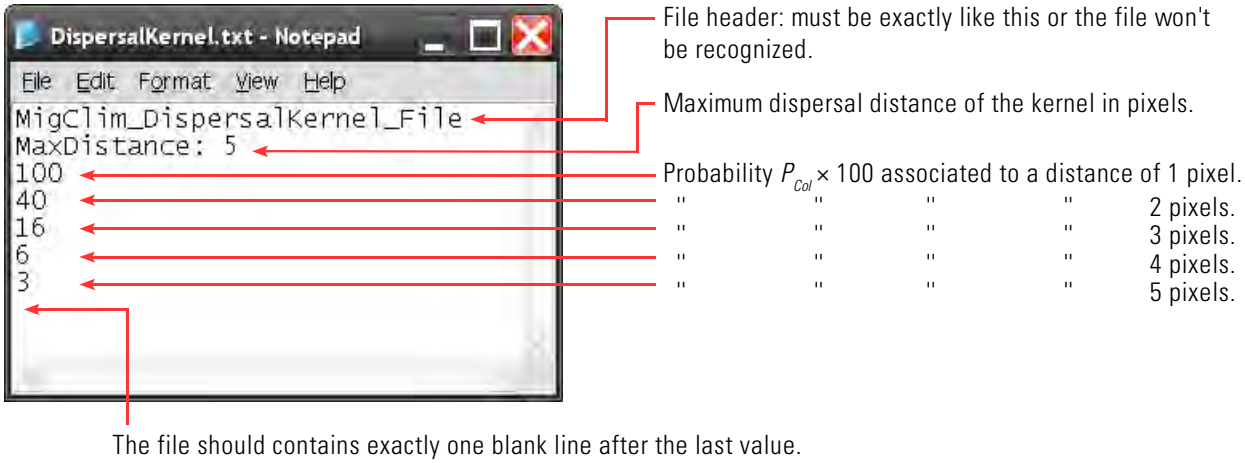
Step 2.1: Dispersal kernel

The second form of the MigClim wizard will ask you to enter values for dispersal distances and related probabilities. What is important to understand at this stage is that MigClim considers two kind of dispersal processes: ***short distance dispersal***, which is the normal dispersal mode and **l*ong distance dispersal (LDD)***, which represents very rare events where seeds are dispersed by non-usual means and travel longer distances than usual. To illustrate this with a real example, think about an anemochorous species, i.e. a species dispersing its seeds by wind. The short dispersal distance for this species represents the usual distance that seeds are dispersed under normal wind conditions, say for instance 50 meters. Long distance dispersal on the other hand corresponds to rare events where a seed will be dispersed over a much longer distance, say up to 1000 meters or 10 kilometers for instance, through an exceptional strong wind, or if it gets attached to an animal that will transport it.

For computing efficiency reasons, the approach taken in MigClim is to simulate these two kind of dispersal as separate processes. For short distance dispersal, you will be asked to indicate a maximum dispersal distance, as well as the probability of P_{Col} to be associated with each distance. For long distance dispersal (LDD), you need to give a minimum and maximum distance (in pixels) and a probability of generating an LDD event.



Input values for short distance can be entered manually in the form or can be loaded from a text file using the ***load from file*** button. This file must have the following structure:



The example above is the "DispersalKernel.txt" file that is part of the test data, but you can use any other self-made file that will have the appropriate structure.

Step 2.2: Entering data manually

First, you have to indicate the maximum short distance dispersal distance for your species in pixel units. In the case of our test data, the pixel resolution is 50 meters (note that the resolution of the grids you are using is indicated to you on the form, in the second line).

For this test simulation, we will consider that our species has a short distance dispersal of 250 meters, which corresponds to 5 pixels (5 × 50=250).

Enter a value of "5" in the ***maximum "short distance" dispersal distance*** box and you will see that the two fields located below (***distance in pixels*** and ***colonization probability***) will become filled automatically. The ***distance in pixels*** field indicates the distance from a source pixel, and the ***colonization probability*** the value of P_{Disp} corresponding to this distance. As can see, the default ***colonization probability*** has been set to 100 for all distances. This means that the probability of colonization is not dependent on distance from the source pixel, and is basically of 100% for all sink pixels within a 5 pixel distance from a source pixel.

Although this default calibration can be useful if you don't want or cannot calibrate values of P_{Disp} more precisely, it is not very realistic as we know that the value of P_{Disp} generally decreases with distance from the source pixel (i.e. the more distance between a source and sink pixel, the less likely for the sink pixel to become colonized). So what we want to do, is to change the value of P_{Disp} so that they decrease with increasing distance from the source pixel.

To do this, select one of the entries in the ***distance in pixels*** field by clicking on it (it will turn grey), then enter a value between 0 and 100 (this value corresponds to P_{Disp} × 100) in the field to the left and press ***Enter>>***. You can enter decimal values. For this test simulation, we set the colonization probabilities as indicated on the figure above.

Once you finished to enter your dispersal kernel, you can save it by clicking on ***Save to File***. This will save your dispersal kernel as a text file in the locations and with the name that you specify. Once you saved your kernel, you can use re-use it easily with the ***Load from File*** button. Using this button, you can also try to load the dispersal kernel that is provided with the test data. The file is named "DispersalKernel.txt" and you will find it in your "C:\MigClim\TestData" folder (or any other location

where you have copied the test data). The "DispersalKernel.txt" file contains the same kernel as shown in the figure above.

*Note: What is the **Colonization probability correction factor for pixel size compensation**?*
This is an optional parameter that allows to modulate the values of P_{Disp} to accommodate for simulations run with different pixel sizes. You can read more on this parameter at the end of this manual ("Part 4: Additional options available in MigClim") but for now we will just leave it at its default value of 1.

Step 2.3: Long distance dispersal (LDD)

Long distance dispersal (LDD) is an optional parameter in MigClim. If you wish to enable it, you should check the ***Enable long distance dispersal*** box, and this is what we will do in this test simulation. Once you have the option enabled, you have to provide the following information:

Minimum LDD distance: The minimum distance (in pixel units) for LDD events.

Maximum LDD distance: The maximum distance (in pixel units) for LDD events.

LDD event frequency: The frequency with which LDD events will be generated. For instance, a value of 0.01 means that and LDD event will be generated in 1 time out of 100 (so, if you have 100 pixels that are mature, on average, 1 of them will produce a LDD event during a given dispersal event).

For our test simulation, we will set the minimum LDD distance to "6" pixels, the maximum LDD distance to "40" pixels and the frequency to "0.01" (= 1%, that's possibly a rather high value).

Important: the minimum LDD dispersal distance has to be larger than the maximum short dispersal distance.

Note: the probability of an LDD event to reach any pixel within the defined distance range [Minimum LDD distance – Maximum LDD distance] is independent of that distance, i.e. the probability remains constant.

Note: implementing LDD dispersal is optional. Simulations can also be run without it.

Note: LDD events are not stopped by barriers.

Step 2.4: Moving on to the next form of the wizard

Once you have entered your dispersal parameter values, you may click ***Next >>*** and proceed to the next form.

Step 3: Increase of pixel reproductive potential over time

Form number 3 deals with the ***increase in reproductive potential over time*** (parameter P_{Mat}). MigClim allows you to specify how the reproductive potential of a colonized pixel increases over time (starting when the pixel becomes colonized). This can be useful if you are modeling a species that needs more than one year to start producing seeds, such as a tree for instance. Furthermore, this parameter also allows you to reflect the fact that a population (i.e. a pixel) will usually have an increase of its reproductive potential over time because the number of individuals within the pixel will increase. In other

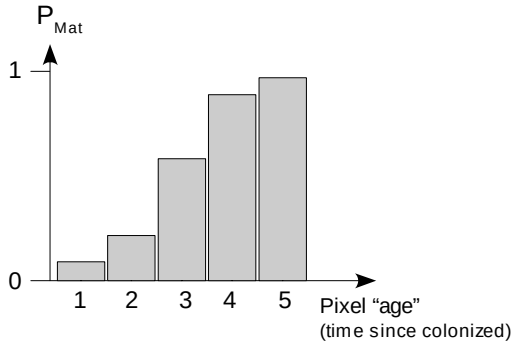
words, the reproductive potential of a given pixel (representing a population) will not immediately reach 100%.

In MigClim, the increase in reproductive potential over time is represented by the parameter P_{mat} , which can take values between 0 and 1:

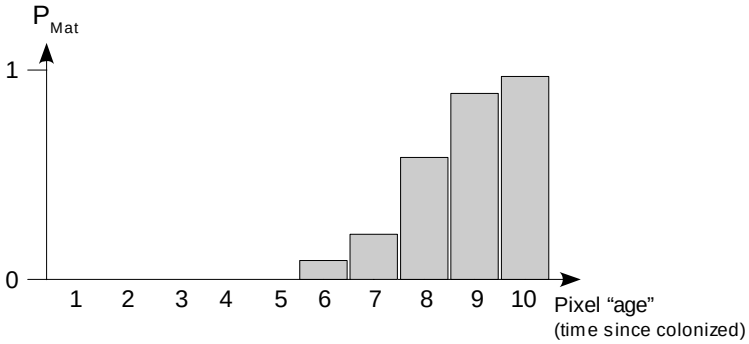
→ 0 = 0% of reproductive potentially.

→ 1 = 100% of reproductive potential = full reproductive potential).

To give you an example, here are two examples of how increase in reproductive potential over time might look:



Example 1. Here the species starts to have some reproductive potential since the first dispersal event (i.e. usually 1 dispersal event = 1 year) after the pixel became colonized (i.e. pixel age = 1). The value of P_{Mat} then increases in a sigmoid fashion to reach its maximum at the age of 5. From the age of 5 on, the value of P_{Mat} will always remain at its maximum of 1



Example 2. In this case, the species will not be able to reproduce before it has reached an age of 6 dispersal events (i.e. generally 1 dispersal event = 1 year).
From the age of 6, the value of P_{Mat} then increases in a sigmoid fashion to reach its maximum at the age of 10. From the age of 10 on, the value of P_{Mat} will always remain at its maximum of 1

Important: While the value of P_{Mat} are in the range [0:1], they have to be in the range [0:100] when you enter them in the MigClim form. In other words the values you enter in the form are $P_{Mat} \times 100$.

Step 3.1: Entering information for the reproductive potential kernel:

In order to define your reproductive potential kernel, you first need to enter information about the time of ***Initial maturity*** and ***Full maturity*** of your species.

Initial maturity: This is the age (in dispersal event units) when a pixel starts to become mature and thus has a value of $P_{Mat} > 0$.

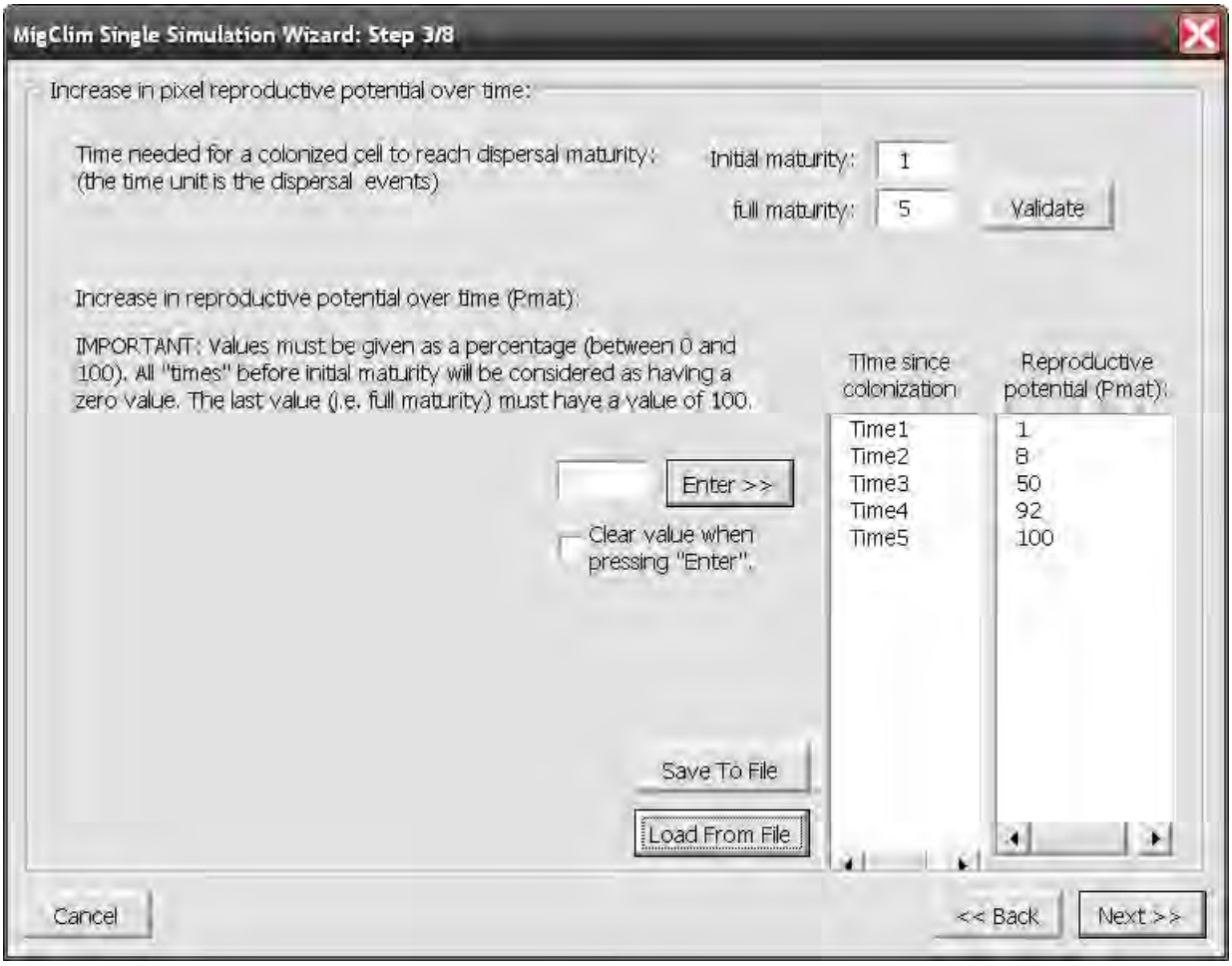
Full maturity: This is the time when a pixel has reached full maturity ($P_{Mat} = 1$).

In our test simulation, we will set our increase in reproductive potential so that it starts at the age of one and reaches 100% at the age of five. To set this, enter "1" in the ***Initial maturity*** field and "5" in the ***Full maturity*** field. Note that you can only enter entire numbers into these fields.

Press the ***validate*** button and you will see two fields of the form, ***time since colonization*** and

reproductive potential, will become populated. The values of "Time1" to "Time4" receive a value of "UnDefined", while the last, "Time5" receives a value of 100. While the value of "Time5" is correct (remember that the last value must always be 100), the values for "Time1" to "Time4" obviously have to be changed. For this test simulation we will enter the values just as in the figure below: 1, 8, 50 and 92.

To update values in the ***reproductive potential*** field, proceed just as you did for the dispersal kernel: select a value of ***time since colonization***, enter a value in the box to the left and press ***Enter >>***.



Once you finished to enter your reproductive potential kernel, you can save it by clicking on ***Save To File***. This will save your kernel as a text file in the locations and with the name that you specify. Once the kernel is saved, you can use re-use it easily with the ***Load From File*** button. Using this button, you can also try to load the increase in reproductive potential kernel that is provided with the test data. The file is named "ReproductivePotentialKernel.txt" and you will find it in your "C:\MigClim\TestData" folder (or any other location where you have copied the test data). The "ReproductivePotentialKernel.txt" file contains the same kernel as shown in the figure above. The structure of a file containing a reproductive potential kernel is explained earlier in this user guide, in the section where the test data is described.

Important: Once pixels have reached their full maturity, their value of P_{Mat} remains always equal to 1 (unless they become decolonized, obviously).

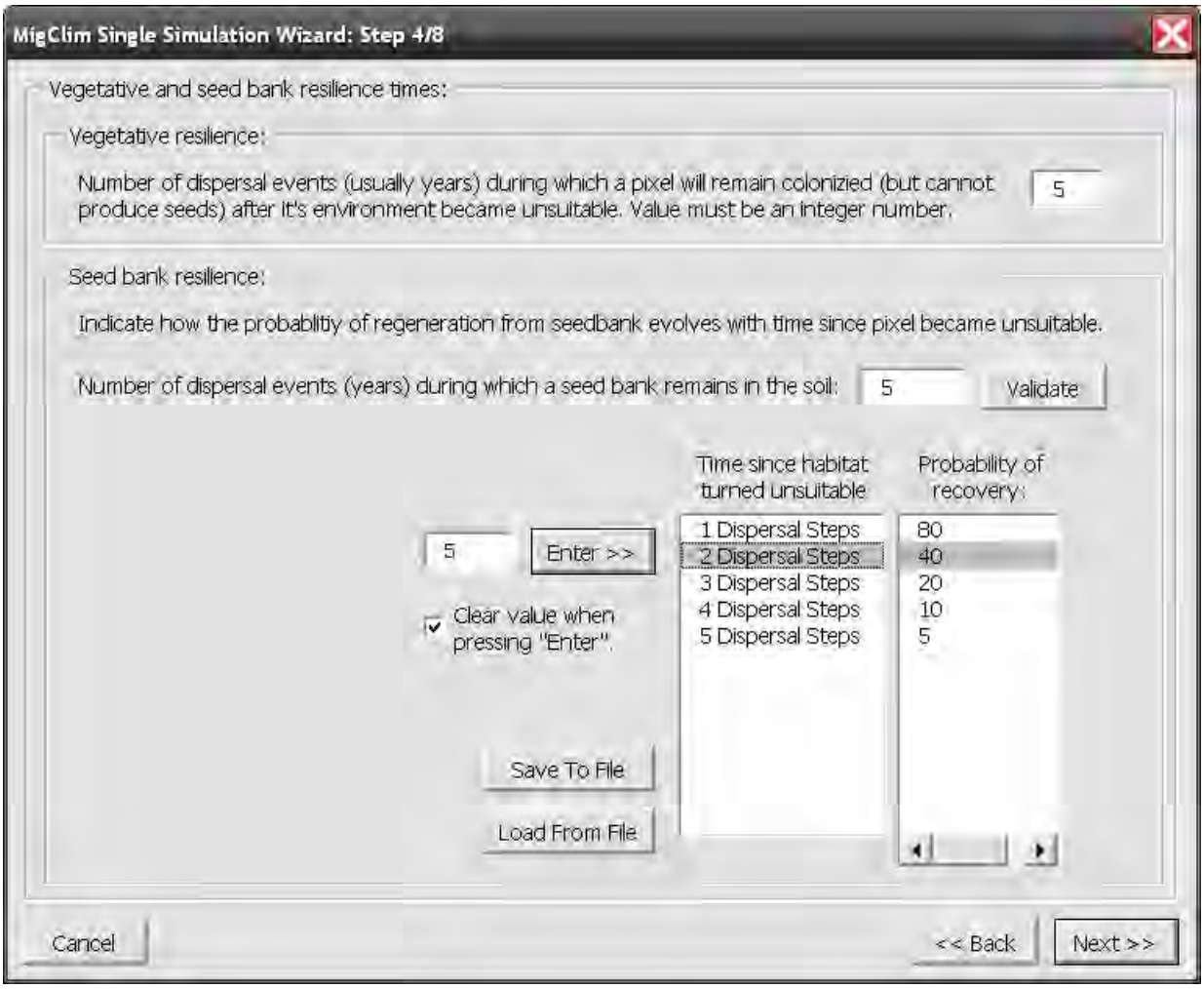
Important: In MigClim, time is measured in terms of dispersal events. Usually one dispersal event should be set so that it equals to one year.

Step 3.2: Move on to the next form

Did you indicate the initial and full maturity values and did you give a P_{Mat} values to each time between the initial and full maturity? Well done, you're ready to proceed to the next form! Hit ***Next >>***.

Step 4: Resilience to environmental change: Vegetative and seed bank resilience time

Form number 4 allows us to specify two further parameters for our species: ***vegetative resilience*** and ***seed bank resilience*** times. These are optional parameters that enable a species to remain in a pixel, either in a vegetative resilient form (i.e. unable to reproduce nor grow) or as a seed bank in the soil, even if the habitat in this pixel has turned unfavorable. If you are not familiar with the concepts of vegetative resilience and seed bank resilience in MigClim, you may want to read again the detailed explanations given on pages 9-12 of this user guide.



Step 4.1: Vegetative resilience time

Implementing ***vegetative resilience*** in your simulation couldn't be more straightforward: simply indicate the time during which a pixel can remain in vegetative resilient state and you're done. As always in MigClim, time is measured in dispersal event units (which should, most of the time, be equal to one year). For this test simulation, let's assume our species can remain 5 years in vegetative resilience state. We therefore fill the ***vegetative resilience*** field with the value "5".

Note: vegetative resilience time does only make sense for simulations that do implement environmental change over time (e.g. climate change).

Note: when calibrating the value of vegetative resilience and seed bank resilience, keep in mind the frequency with which you choose to update your habitat suitability maps. For instance, in this test simulation we update the habitat suitability map every 5 years. It would thus make little sense to set a vegetative resilience time that is < 5, because in this case there would be no chance for a pixel to ever recover (because the pixel would leave vegetative resilience before the habitat suitability map was updated).

Step 4.2: Seed bank resilience time

First the bad news: implementing ***seed bank resilience*** will require a little more effort from you. The good news however is that it is a lot of fun, as you will soon discover.

To start with, you must indicate for how long a pixel will keep its seed bank. Here we will set this value to 5 years (well, 5 dispersal events if we want to be formal, but remember that in our test simulation we consider that our species disperses once a year and thus 1 dispersal event equals 1 year).

Once you entered the value "5", press ***Validate*** and let the magic happen... I guess that by now you must know the kind of ("lame" is the word you're looking for) magic I'm talking about...which is simply to see the two fields on the form auto-fill!

By default, the values of ***Probability of recovery*** are all set to zero, which we obviously want to change because if we wanted to have a probability of recovery that is null then we wouldn't implement seed bank resilience in our simulation in the first place.

For each time period since a pixel became decolonized (measured, as always, in "dispersal events"), we must indicate what is the "fitness" of the seed bank. If the conditions for pixel recovery through seed-bank are fulfilled, this fitness value will be used by MigClim as a probability of generating a new population form the pixel. For our test simulation, I suggest that our species has 80% chances of recovering from its seed bank after 1 dispersal events (i.e. 1 year), 40% after 2 dispersal events (i.e. 2 years), and so on (see print screen above) until 5% after 5 dispersal events. These percentages should reflect the fact that a seed-bank is loosing recovery power over time. Once the seed bank resilience time is over, the pixel returns to an "unoccupied" state.

To enter vegetative and seed bank resilience values in the form, you can either fill the form manually, or load this data from a file using the ***Load From File*** button. In this tutorial, we will load our data from the "ResilienceKernel.txt" text file that comes with the test data. A description of this file was given earlier in this tutorial (see the "Test Data Description" section).

Note: If you do not wish to use the vegetative resilience and/or seed bank resilience options, simply leave their values to zero (you can also use either the one or the other).

Step 4.3: Move ahead...

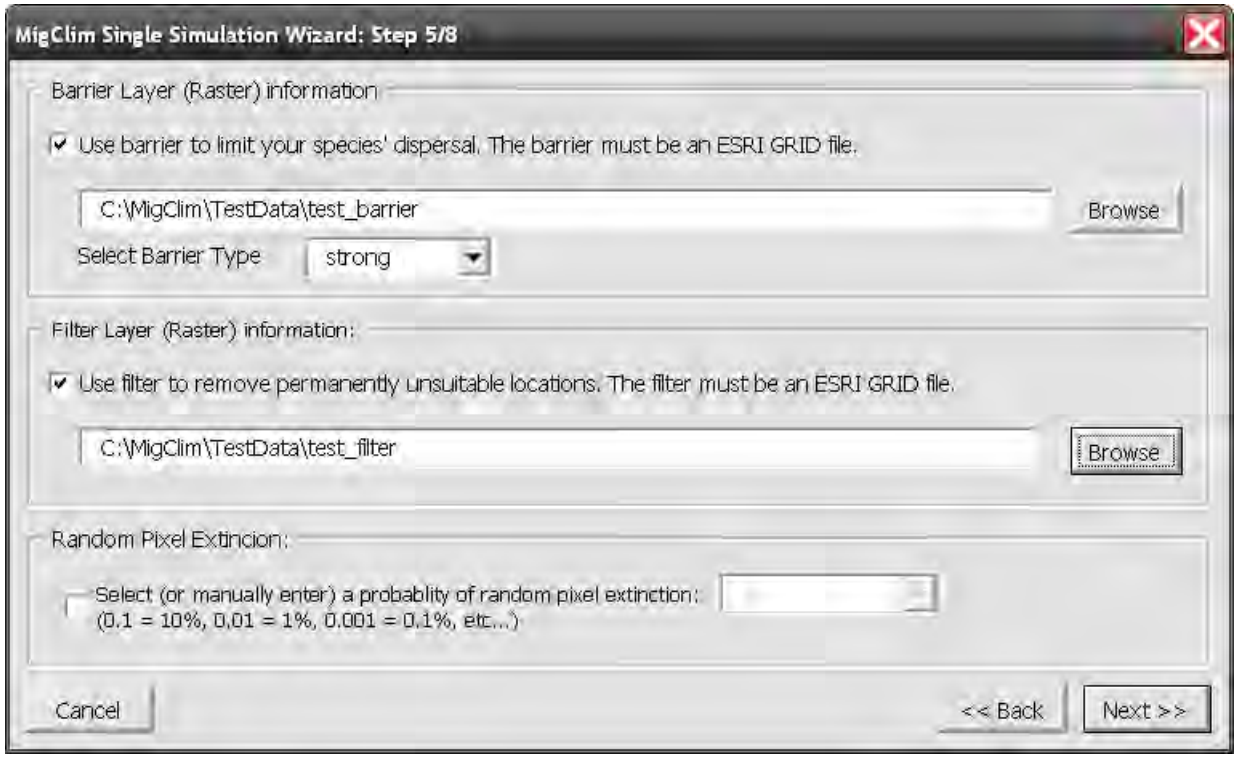
Congratulations, you've calibrated the vegetative resilience and the seed bank resilience parameter for your species. Click ***Next >>*** and you're off to the next form.

Step 5: Landscape fragmentation: Barriers and Filters

In form number 5 you are now given the possibility of specifying a ***barrier*** and/or a ***filter*** layer. These must again be ESRI GRID files. Explanations on barriers and filters are given after the figure.

Important: Barrier and filter girds must be in the ESRI GRID format and have exactly the same extent and pixel size than all other grids (i.e. initial distribution and habitat suitability maps) that you are using in your simulation.

Note: Barriers and filters are optional parameters, simulation can also be run without them.

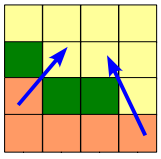


Step 5.1: Barrier layer

A ***barrier layer*** is a raster that represents barriers to dispersal processes, i.e. pixels through which dispersal is impeded. For instance, a forest pixel could be a barrier to a strictly grassland plant, or highways might be barriers for species that are dispersed by ants (or maybe I do not know about and ecology).

Barriers pixels must have a value of "1", while non-barrier pixels must have a value of either "0" or "NoData". As already mentioned earlier, I would always try to avoid "NoData" values as much as possible and stick to "0" for indicating non-barrier pixels.

There are two kind of barriers that can be implemented in MigClim: ***strong*** barriers and ***weak*** barriers. Basically, the difference is that ***strong*** barriers are more restrictive than ***weak*** barriers. The figure below illustrates the differences between these two kind of barriers:



In this figure, green pixels represent barriers, orange pixels represent colonized pixels and yellow pixels are empty but suitable pixels. The blue arrow indicate dispersal movements that are allowed if we use a weak barrier, but impeached if strong barriers are used. Strong barriers are thus more restrictive than weak barriers.

For our test simulation, we will use the barrier raster file called "test_barrier". Click ***Browse*** to locate this file on your computer or copy-paste its location in the relevant field. We will choose to implement our barrier as a ***strong barrier***, which is the default option of the drop-down menu.

Step 5.2: Filter layer

A ***filter layer*** is a raster that represents locations that remain permanently unsuitable during the entire duration of the simulation. For instance, an urban area could be such a permanently unsuitable area. A filter must be given as an ESRI GRID raster with values of 0 = filter pixels and values of 1 = non-filter pixels. For the purpose of this tutorial, we will use the "test_filter" grid file that came with this userguide's test data. You can browse to this file using the ***Browse*** buttons.

Note: make sure not to be confused between barrier and filter layers. In barrier layers, those pixels that are part of the barrier have a value of "1," while in filter layer, those pixels part of the filter have a value of "0".

Step 5.3: Random Pixel Extinction

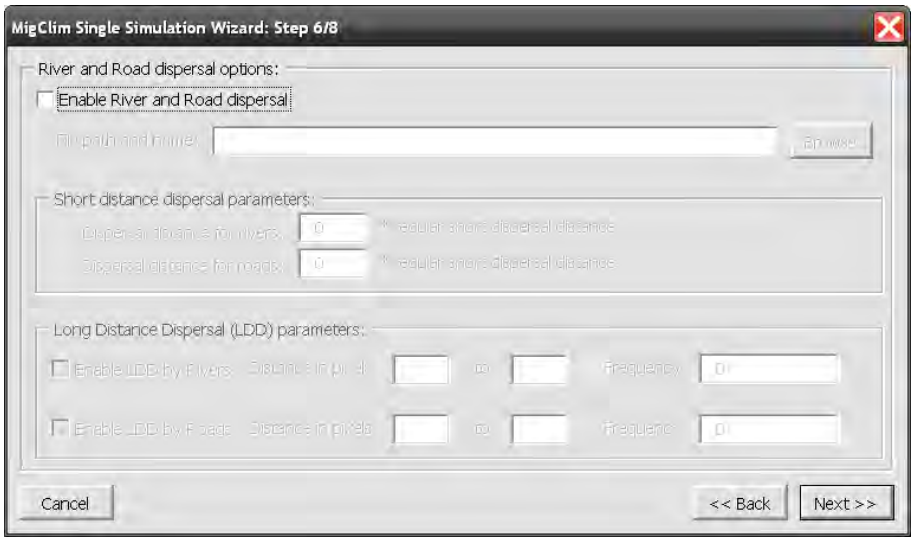
Form number 5 also offers you the opportunity to set a ***Random Pixel Extinction*** rate. Enabling this option (i.e. ticking the box) will randomly and with the indicated probability, reset some occupied pixels to unoccupied. Note that you can either pick one of the probabilities from the drop-down list, or enter a probability yourself. The probability value must be between 0 and 1 (0=0%, 1=100%).

In this test simulation, we will not use the *Random Pixel Extinction o*ption, because this adds a supplementary random component to the simulation and makes the outputs more difficult to interpret. So leave the box unchecked in this section and click ***Next >>*** to cruise along to the next form.

Step 6: River and Road dispersal

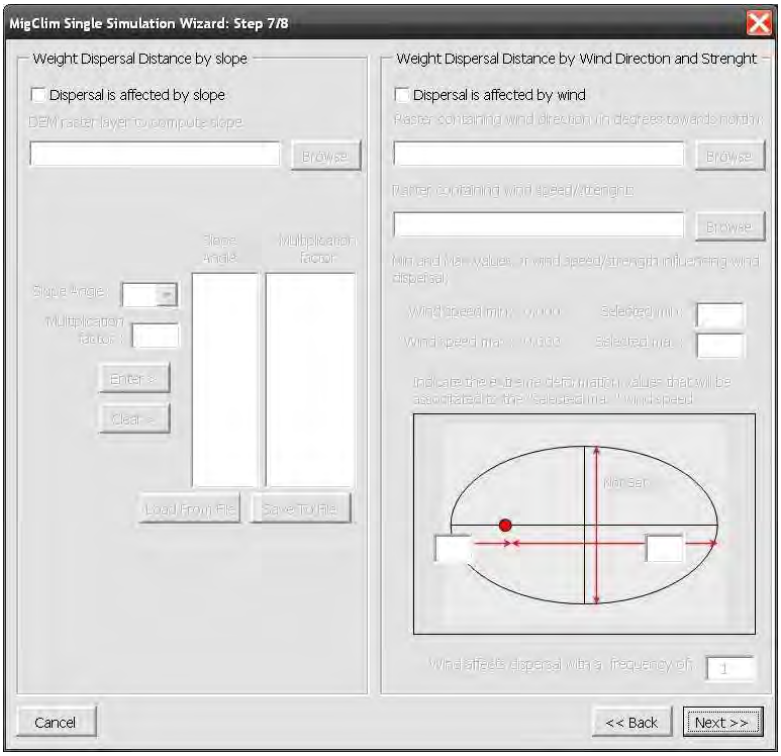
Form number 6 is home of the ***river and road dispersal*** option. This option allows to implement differential dispersal distances along river and/or road features. For instance, a species might disperse longer distances along rivers or road features than its normal dispersal distance.

In this exercise however, we are not going to use this option, so just click **Next >>** and move along to the next form. If you are interested to read more about the *river and road dispersal* option, more informations are given later in this user guide (see "Part 4: Additional option available in MigClim").



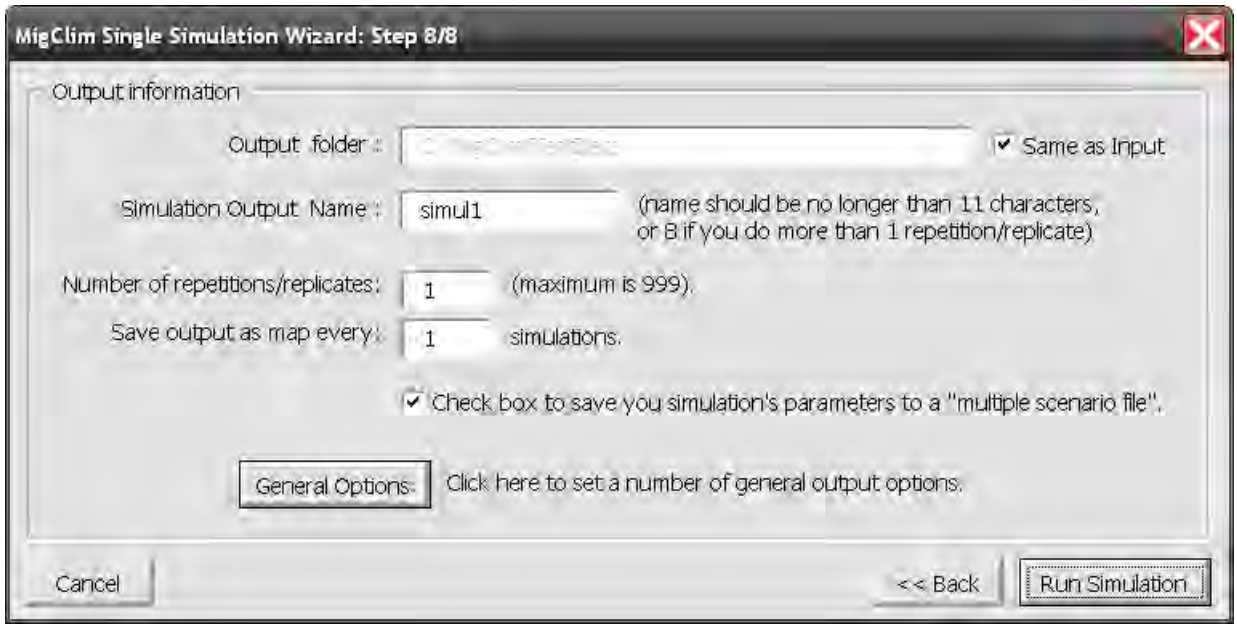
Step 7: Slope and Wind parameters

Form number 7 is another purely optional form. Here you can choose to locally weight the dispersal distance through ***slope*** or ***wind speed and direction***. Again, we won't use these options in this tutorial, so simply skip this step by clicking ***Next >>***. Yes, we're already done with this form (starting to like this tutorial?). More information on this option are given in "Part 4: Additional option available in MigClim".



Step 8: Output Information

This is it: the last form of the wizard, and it's even a really easy one to complete. So let's see what we've got here.



Step 8.1: Output name and location

First of all, we are asked for an ***Output Folder***. This is where all the outputs from our simulation will be saved, so make sure you have full write access to the folder you indicate or there will be troubles down the road.

For the purpose of this tutorial, we shall chose to save our data in the same folder as the input data. To do so, we can simply tick the ***same as input*** box.

Important: the output folder must already contain at least one ESRI GRID file. If not then this directory won't be recognized as a valid workspace and your simulation won't run.

Next we're asked for an ***Output Raster Name***. This is simply the name that will be given to your simulation and that will be used to name the output grids (rasters). Note that this name cannot exceed 8 or 11 characters (depending if you're doing one or more repetitions/replicates of your simulation), and is not allowed to start with a numeric character, but can contain numeric characters within or at the end (e.g., "1simul" is not acceptable, but "simul1" is).

Step 8.2: Number of repetitions/replicates

To continue filling the form, we must now indicate the ***Number of repetitions*** and the ***Save output as map every ... simulation*** fields. The idea behind these parameters is simple: since MigClim simulations generally involve some randomness (this actually depends on the parameter values that you have entered), it can be interesting to repeat several times a same simulation and look at the variability due to random processes.

The number of times the simulation should be repeated has to be indicated in the ***Number of repetitions*** field.

The ***Save output as map every ... simulation*** field allows you to specify with which frequency you

would like to save your output as a map (raster). Let me clarify this: in all cases, the MigClim output is always saved to a text file, giving you for instance the number of pixels that are colonized or lost at each dispersal step of the simulation. While having the output as a map too is certainly interesting (because it is spatially explicit, which is not the case of the text file), it might not be necessary to have 100 output maps if you do 100 repetitions for instance. In such a case, you might want to export the output as a raster only every 10th simulation in example. You would hence indicate ***Save output as map every 10 simulation***.

For this tutorial, we will run our simulation only once, and we will thus save our output map every 1 simulation. You may later try to run the same simulation but this time indicating to repeat it say 5 times to look at the variability of your output.

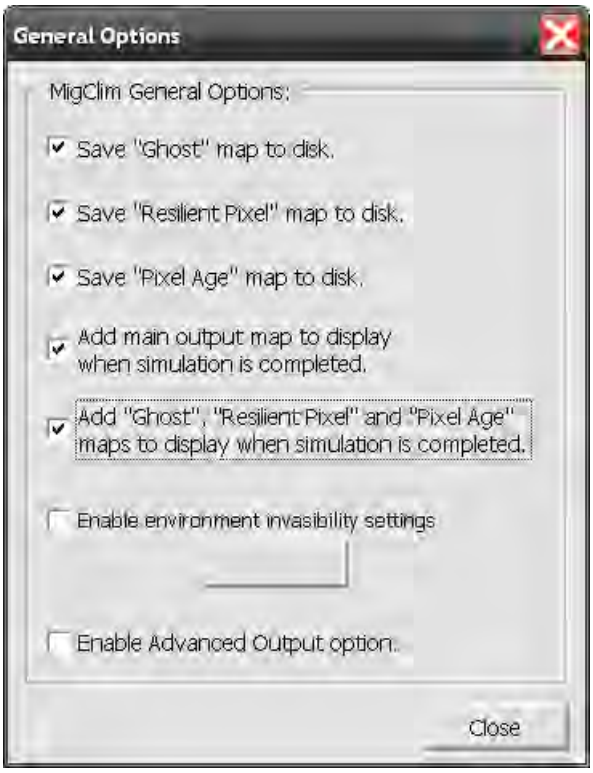
Step 8.3: Additional output options

At this stage, you have probably noticed a check-box named ***Check box to save your simulation's parameters to a "multiple scenario file"***. If you activate this option, MigClim will create a text file, named "MigClim_Scenario_Log.txt" and located in your output directory, that contains the instructions to run your simulation (with exactly the parameters you specified) in batch mode. This is very handy, because it means that you can start your simulation in just a few clicks rather than having to go through the whole wizard all over again.

Note: If you chose to save your parameters to a "multiple scenario file" but you have not saved your dispersal kernel / reproductive potential kernel / resilience time kernel to a text file, then you will be prompted to do so.

Note: In fact MigClim will always generate a "multiple scenario file" called "MigClim_Scenario_Log.txt," even if you didn't check the above mentioned box. If you leave the box unchecked and did not save some of the files for "Dispersal Kernel," "Reproductive Potential Kernel" or "Resilience Kernel," then MigClim will generate these files automatically and give them a default name (starting with "zzTemporary").

Finally, you are given the opportunity to modify the ***General Options*** of MigClim before starting the simulation. You can of course also modify the general options through the ***Options*** menu in the MigClim toolbar, but if you have forgotten to do it, now is the last call before take-off. For our test simulation we shall check all but the last two of these options, as illustrated in the figure below:



Note: since the aim of this user guide is to explain MigClim, I have here chosen to activate most of the options of the form. If you are using it for yourself, you might want to deactivate some of them. For instance, you might not necessarily be interested in keeping the "Ghost" map of the "Pixel Age" map. More explanations on what all these options exactly do is given earlier in this user guide.

Step 8.4: Believe it or not... we're done :-)

Entered everything? Go ahead and click ***Run Simulation*** and....hopefully, the MigClim simulation should start! You can follow the simulation's progress in the status bar (bottom left) of your ArcMap window. Running this simulation on an laptop computer (1.8 GHz Pentium M) took about one minute, but your millage might vary depending on the kind of machine you're working on.

Once the simulation has completed, a message displays, indicating the total time of your simulation (this information will also appear in the output text file so you don't need to write it down).

In the next section of the tutorial we will look at the different outputs produced by MigClim and see how to interpret them.

Interpretating MigClim's Output

Hopefully your simulation will have gone well and you should now find the following files in your output folder "C:\MigClim\TestData", or any other folder that you have indicated as output folder.

4 new rasters:

simul1: the main output, showing the distribution of the test species at the end of the simulation.

simul1go: the "Ghost" map that has kept track of all pixels that have ever been colonized.

simul1re: basically the same output as "simul1" but also showing "vegetative resilient" pixels.

simul1ti: a map that shows the age (in dispersal event units) since a given pixel was colonized. Note that in our case one dispersal event equals to one year.

3 new text files:

MigClim_FullOutput.txt: a file that contains the complete output of your MigClim simulation.

MigClim_SummaryOutput.txt: a file that contains a summary of the simulation's output.

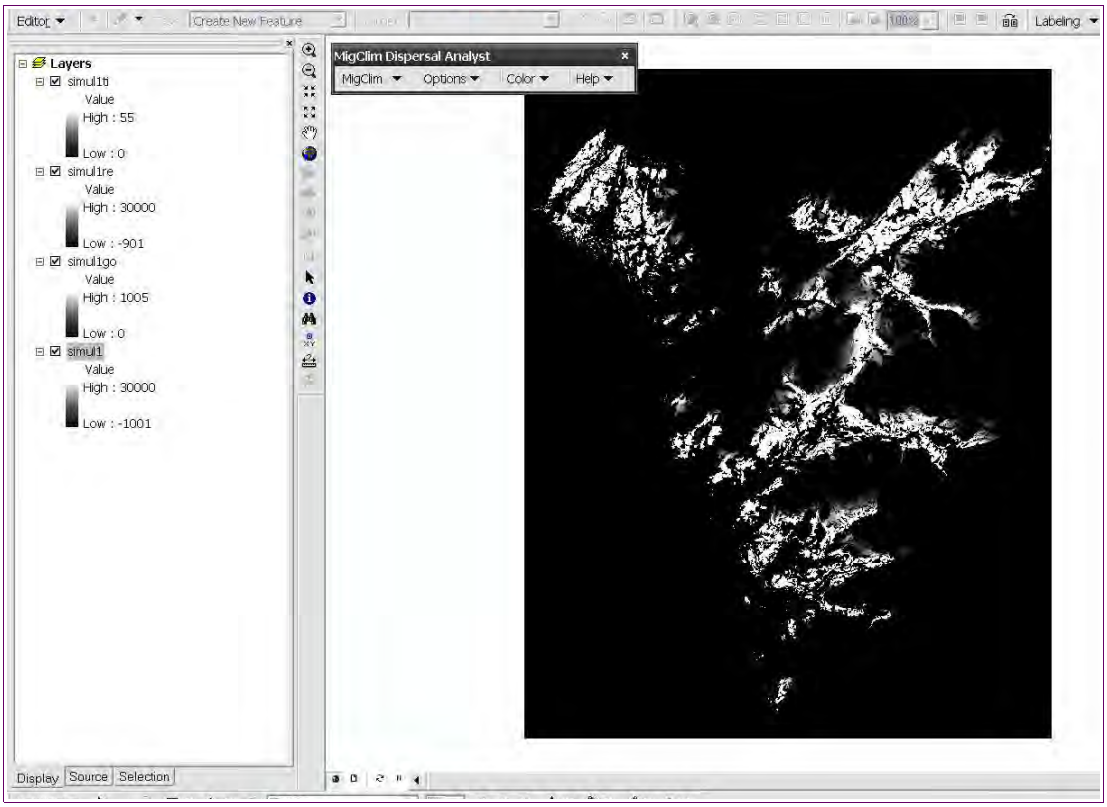
MigClim_Scenario_Log.txt: a file that contains all the information of the simulation you just ran and that can be used to launch the same simulation again in just a few clicks using the "multiple simulation" option of MigClim, a sort of batch mode (but if you do so make sure that you changed the output name of your simulation or delete the existing output rasters. You can keep the output text files though, the new results will simply be appended within the existing files).

Important: In MigClim, text file outputs are not overwritten. Instead, if you run a new simulation, the results from this simulation will simply be appended at the end of the existing text file. Note that the name of the three output files are always the same and cannot be changed. For raster/maps outputs, the story is a little different: raster output have a name that is based on your choice of "simulation output name". Raster outputs cannot be overwritten and if you attempt to start a simulation that has the same output name as an existing raster, the simulation will fail (actually MigClim is supposed to tell you that something is wrong with your simulation and impede you from even starting it).

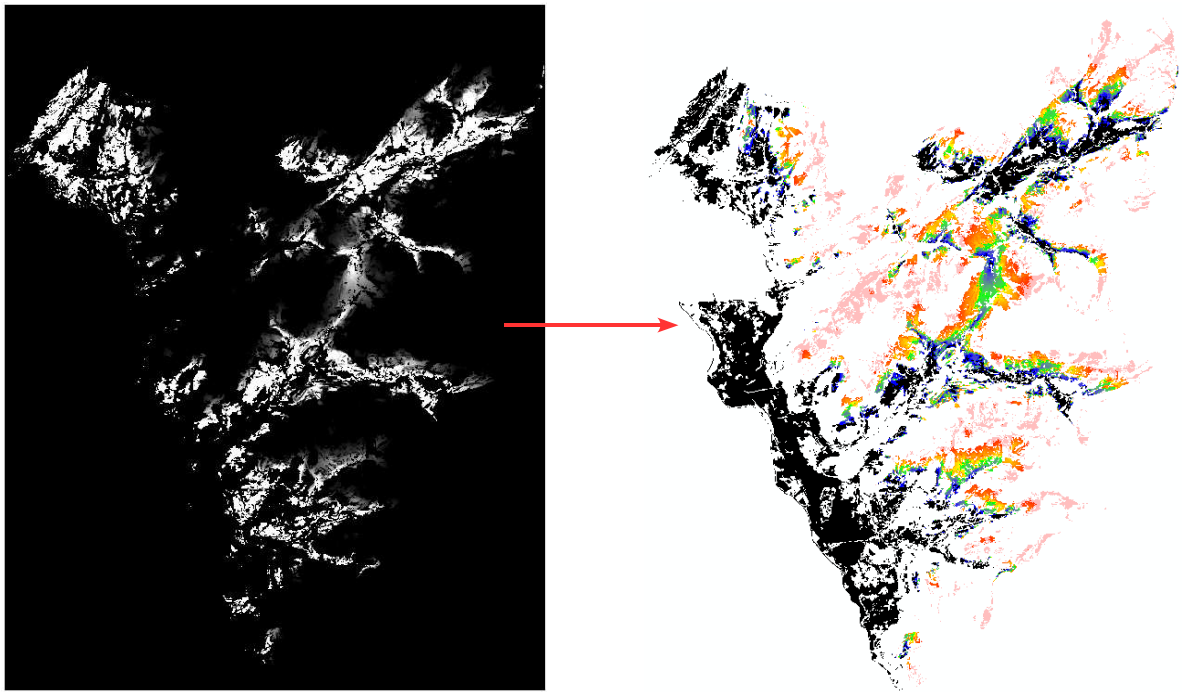
If you wish to verify your results, you can compare them to those you will find in a folder called "ExampleOutput" that came with the test data. Be aware however that, because there are random processes involved in the test simulation we just did, your results will not be exactly the same to those you'll find in the "ExampleOutput" folder. They should however, be fairly similar.

Since we did check the option to add output maps to the display once the simulation has completed, you should find the four following layers in your ArcMap application: **simul1**, **simul1go, simul1re** and **simul1ti.**

As you will notice, all those grids (maps) have received a default color schemes that make them not only artistically unappealing, but also difficult to read and interpret.



To get the whole thing a bit more colorful and easier to read, use the ***Color*** > ***Color All Visible Rasters*** button of the MigClim toolbar. Doesn't that look much better?



Interpretation of raster (grids) outputs

Alright, that was the easy bit of this section! We shall now look at the different maps (rasters) and understand what they mean.

simul1 (main output):

As already mentioned, this is the main map output from MIGCLIM and it has been given exactly the name you specified as "Simulation output name". The pixel values of this raster have to be interpreted as follows:

Pixel value	Signification
0	Pixels that have never been colonized and are unsuitable habitat at the end of the simulation. There are several reasons why pixels can remain non-colonized: - They have remained unsuitable during the entire simulation. - They are part of a barrier or a filter. - Their colonization has failed for dispersal-limitation related reasons.
1	Pixels that belong to the species' initial distribution and that remain suitable at the end of the simulation.
1 < value < 30'000	Positive values greater than 1 but smaller then 30'000 represent pixels that have been colonized during the simulation and that remain suitable and occupied at the end of the simulation. The value of the pixel allows to determine the dispersal event during which it was colonized using the following code: each environmental change event is given a value of 100 and each dispersal event a value of 1. Here are some examples: 101 = 1st dispersal event of 1st environmental change event (1 × 1 "+ 1 × 100 = 101). 102 = 2nd dispersal event of 1st environmental change event (2 × 1 "+ 1 × 100 = 102). 504 = 4th dispersal event of 5th environmental change event (4 × 1 "+ 5 × 100 = 504). 1003 = 3rd dispersal event of 10th environmental change event (3 × 1 "+ 10 × 100 = 101).
30'000	Pixels that are potentially suitable (i.e. habitat is favorable) but that were not colonized due to dispersal limitations. These pixels represent the difference between the unlimited dispersal scenario and the current simulation.
Value < 1	Negative values indicate pixels that were once colonized but have become decolonized, either because their habitat has turned unsuitable or because they have undergone a user-implemented extinction event (i.e. P_{Ext} parameter). Their absolute value allows determining the exact time when they have been decolonized using the same code as explained just above.

Note: If you are running a simulation without environmental change, then there will be no negative values either (actually there can be some negative values, but only if the simulation implements pixels extinctions through the P_{Ext} parameter).

simul1go ("Ghost" raster/map):

This raster has the two letters "go" appended to the name that was specified as "Simulation output name". Those stand for "Ghost", because this raster shows all pixels that were ever colonized during a simulation, even if those pixel eventually became extinct (e.g. because of environmental/climate change). In other words, this raster allows you to see the ghost of some pixels that were once colonized but have since become extinct either because their habitat has turned unsuitable or because the user has chosen

to implement random pixel extinction through the P_{Ext} parameter.

The pixel values in the "*Ghost*" raster indicate the dispersal event during which a pixel was colonized, using the same code that was explained in the table just above.

Important: While the "Ghost" raster keeps track of all pixels that have ever been colonized during a simulation, it only records the time when a pixel was first colonized. This means that if a pixel gets colonized several times, only the time of its first colonization will be recorded in the "Ghost" raster.

Note: If you are running a simulation without environmental change, then the "ghost" raster gives little or no extra information as compared to the main raster. The only information the "ghost" raster might give in this case is an indication of which pixels have been decolonized through the P_{Ext} parameter. Those pixels will have the value of when they first got colonized in the ghost raster, but a negative or a different positive value (if they have been re-colonized) in the main raster.

simul1re ("Resilience" raster/map):

This raster has the two letters "re" appended to the name that was specified as output. Those stand for "resilience" and this is because this raster also includes those pixels that are in a vegetative resilience state (in the main raster, the vegetative resilient pixel appear as extinct).

This raster is in fact exactly the same as the main output raster (i.e. "simul1") with the only difference that it also shows pixels that are in a vegetative resilience state. The pixel values are interpreted using the same code than the one that was explained just above for the main raster output.

Note: In simulations run without environmental change, this output is not produced because it would be exactly the same as "simul1".

simul1ti ("Time" raster/map):

This raster has the two letters "ti" appended to the name that was specified as output. Those stand for "time" and this is because this raster indicates the "age" of each pixel, i.e. the time since it was colonized. If a pixel was colonized more than once, then the "Time" raster value indicates the age of the pixel since it was last colonized. If the pixel has spent some time in vegetative resilience stage (reminder: in this state a pixel does not "grow" anymore and thus its age does not increase anymore), then this period of time will, logically, not appear in the pixel's "age".

To give you a few examples, a pixel value of 1 means that this this pixel was colonized during the last dispersal event of the last environmental change step. A value of 50 on the other hand indicates that the pixel has been colonized since 50 dispersal events (=50 years in our case). For our simulation, this means that it has been colonized during the very first dispersal event/step of the simulation because out test simulation was run over 50 years (5 dispersal events × 10 environmental change events = 50 dispersal events = 50 years).

Since our simulation was carried-out over a 50 year time period, then you must (or should) be wondering how comes that some pixels have values of 55 and 255. Well, those are somewhat special values: 55 indicates pixels that belong to the species initial distribution. If you remember, and I'm sure you do but I'll give you a refresher nevertheless, I explained somewhere earlier in this tutorial that pixels belonging to the initial distribution are considered as fully mature. Since we specified that pixels reach their full maturity after 5 years, those pixels received an "age" of 5 at the start of the simulation. Add the 50 years of the simulation to those initial 5 and that's how we end-up with a pixel value of 55.

The value of 255 is a special value indicating that the pixel is in a "seed bank" resilient state. Note that "vegetative resilient" pixels have an "age" that is the same as when they entered this state. E.g., if a pixel is 3 years old when his habitat turns unsuitable and it switches to "vegetative resilient", then it

keeps this age of 3 years until its status is changed.

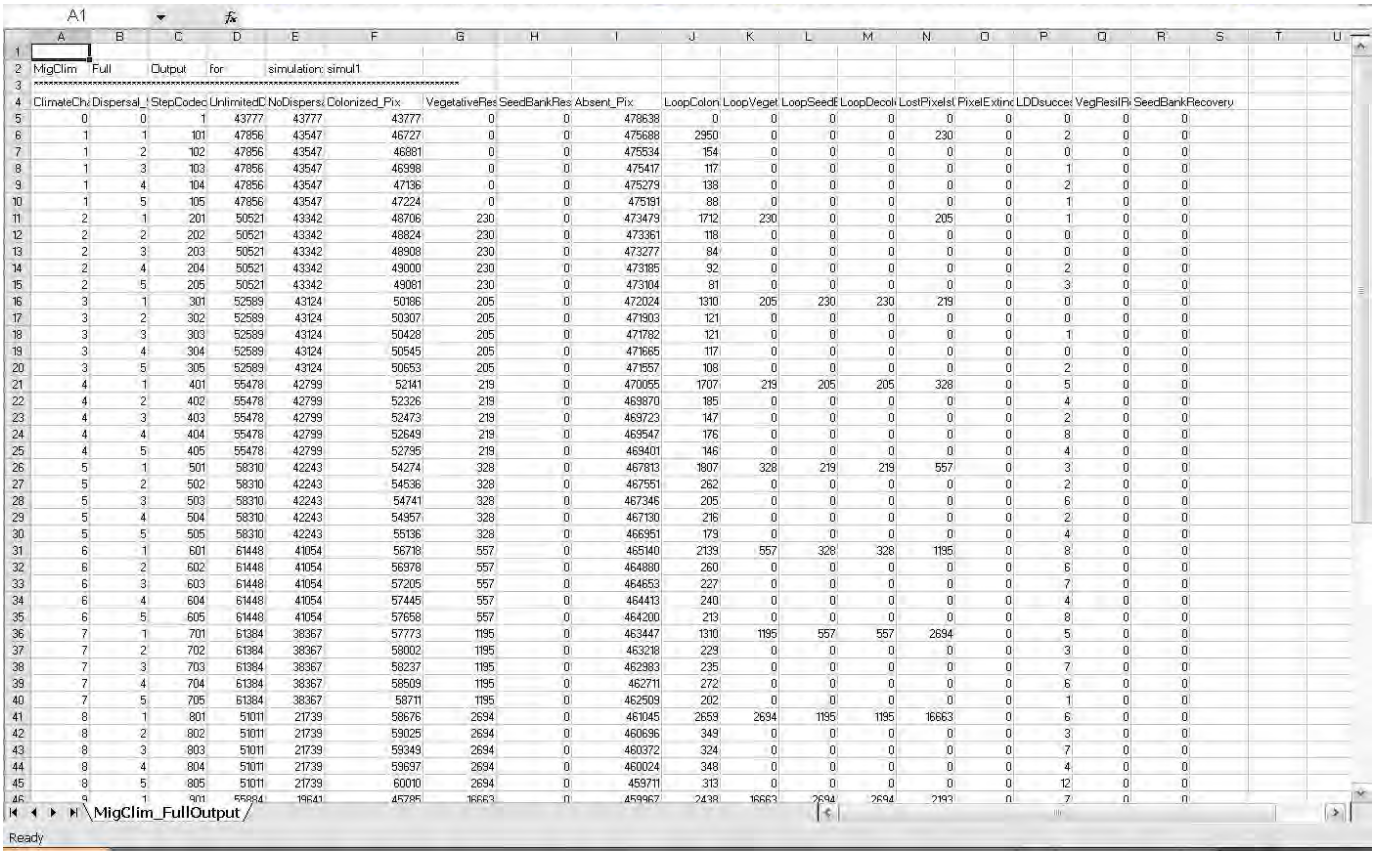
Interpretation of text file outputs

Now that we have looked at all the raster outputs we will delve into the text files outputs. The text file outputs contain more information than the raster files, but the information is no longer spatially explicit.

Important: Well, I guess this is what you call a bad example, but actually in the test simulation there is no pixel that turns unsuitable and then suitable again. As a result, none of the pixel can ever recover through vegetative or seed-bank resilience, so it is normal that your output files will indicate that no pixel has ever recovered.

MigClim_FullOutput.txt:

This is a file that contains the complete output of your MigClim simulation. To interpret this file, I would advise to open it in Microsoft "Excel" or OpenOffice "Calc" using a space (" ") as delimiter. In this way, the data will be nicely displayed in different columns. If you try to open this file as I just suggested, you should have something that looks like this:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1																					
2	MigClim	Full	Output	for	simulation: simul1																
3	**********																				
4	ClimateCha	Dispersal_	StepCodec	UnlimitedD	NoDispersa	Colonized_Pix	VegetativeRes	SeedBankRes	Absent_Pix	LoopColon	LoopVeget	LoopSeedE	LoopDecol	LostPixels(	PixelExtinc	LDDsucces	VegResilFa	SeedBankRecovery			
5	0	0	1	43777	43777	43777	0	0	478638	0	0	0	0	0	0	0	0	0			
6	1	1	101	47856	43547	46727	0	0	475688	2950	0	0	0	230	0	2	0	0			
7	1	2	102	47856	43547	46881	0	0	475534	154	0	0	0	0	0	0	0	0			
8	1	3	103	47856	43547	46998	0	0	475417	117	0	0	0	0	0	1	0	0			
9	1	4	104	47856	43547	47136	0	0	475279	138	0	0	0	0	0	2	0	0			
10	1	5	105	47856	43547	47224	0	0	475191	88	0	0	0	0	0	1	0	0			
11	2	1	201	50521	43342	48706	230	0	473479	1712	230	0	0	205	0	1	0	0			
12	2	2	202	50521	43342	48824	230	0	473361	118	0	0	0	0	0	0	0	0			
13	2	3	203	50521	43342	48908	230	0	473277	84	0	0	0	0	0	0	0	0			
14	2	4	204	50521	43342	49000	230	0	473185	92	0	0	0	0	0	2	0	0			
15	2	5	205	50521	43342	49081	230	0	473104	81	0	0	0	0	0	3	0	0			
16	3	1	301	52589	43124	50186	205	0	472024	1310	205	230	230	219	0	0	0	0			
17	3	2	302	52589	43124	50307	205	0	471903	121	0	0	0	0	0	0	0	0			
18	3	3	303	52589	43124	50428	205	0	471782	121	0	0	0	0	0	1	0	0			
19	3	4	304	52589	43124	50545	205	0	471665	117	0	0	0	0	0	0	0	0			
20	3	5	305	52589	43124	50653	205	0	471557	108	0	0	0	0	0	2	0	0			
21	4	1	401	55478	42799	52141	219	0	470055	1707	219	205	205	328	0	5	0	0			
22	4	2	402	55478	42799	52326	219	0	469870	185	0	0	0	0	0	4	0	0			
23	4	3	403	55478	42799	52473	219	0	469723	147	0	0	0	0	0	2	0	0			
24	4	4	404	55478	42799	52649	219	0	469547	176	0	0	0	0	0	8	0	0			
25	4	5	405	55478	42799	52795	219	0	469401	146	0	0	0	0	0	4	0	0			
26	5	1	501	58310	42243	54274	328	0	467813	1807	328	219	219	557	0	3	0	0			
27	5	2	502	58310	42243	54536	328	0	467551	262	0	0	0	0	0	2	0	0			
28	5	3	503	58310	42243	54741	328	0	467346	205	0	0	0	0	0	6	0	0			
29	5	4	504	58310	42243	54957	328	0	467130	216	0	0	0	0	0	2	0	0			
30	5	5	505	58310	42243	55136	328	0	466951	179	0	0	0	0	0	4	0	0			
31	6	1	601	61448	41054	56718	557	0	465140	2139	557	328	328	1195	0	8	0	0			
32	6	2	602	61448	41054	56978	557	0	464880	260	0	0	0	0	0	6	0	0			
33	6	3	603	61448	41054	57205	557	0	464653	227	0	0	0	0	0	7	0	0			
34	6	4	604	61448	41054	57445	557	0	464413	240	0	0	0	0	0	4	0	0			
35	6	5	605	61448	41054	57658	557	0	464200	213	0	0	0	0	0	8	0	0			
36	7	1	701	61384	38367	57773	1195	0	463447	1310	1195	557	557	2694	0	5	0	0			
37	7	2	702	61384	38367	58002	1195	0	463218	229	0	0	0	0	0	3	0	0			
38	7	3	703	61384	38367	58237	1195	0	462983	235	0	0	0	0	0	7	0	0			
39	7	4	704	61384	38367	58509	1195	0	462711	272	0	0	0	0	0	6	0	0			
40	7	5	705	61384	38367	58711	1195	0	462509	202	0	0	0	0	0	1	0	0			
41	8	1	801	51011	21739	58876	2694	0	461045	2659	2694	1195	1195	16863	0	6	0	0			
42	8	2	802	51011	21739	59025	2694	0	460696	349	0	0	0	0	0	3	0	0			
43	8	3	803	51011	21739	59349	2694	0	460372	324	0	0	0	0	0	7	0	0			
44	8	4	804	51011	21739	59697	2694	0	460024	348	0	0	0	0	0	4	0	0			
45	8	5	805	51011	21739	60010	2694	0	459711	313	0	0	0	0	0	12	0	0			
46	9	1	901	55884	19641	45785	16663	0	459967	2438	16663	2694	2694	2193	0	7	0	0			

The following table will give you a detailed description of what the values of each column mean:

Column Name	Description
EnvironmentalChange_Step	Number of the environmental change event. In the case of our simulation, we have implemented environmental change every 5 dispersal events, therefore you can see that there are always 5 successive lines with the same "EnvironmentalChange_Step" value. *Note: The fist line has a "EnvironmentalChange_Step" value of "0" because this line contains the information regarding the initial distribution of the species, before the simulation was started.*
Dispersal_Step	Number of the dispersal event. In the case of our simulation, we have implemented 5 dispersal events within each environmental change event, therefore you can see that this column always repeats the numbers 1 to 5. *Note: The fist line has a "Dispersal_Step" value of "0" because this line contains the information regarding the initial distribution of the species, before the simulation was started.*
StepCodedValue	This column indicates the "coded" value associated to each dispersal event. The code's explanations were given earlier in this user guide (see the table in the section on raster outputs interpretation). The first line of this column contains the number "1" indicating that this line contains the information about the initial distribution of the species, before the simulation was started.
UnlimitedDispersal	This column contains the number of pixels that would be occupied in the case of the "Unlimited Dispersal" scenario (i.e. a scenario in which we would consider the dispersal of a species as unrestricted). The first line of this column contains the number of pixels of the species' initial distribution. Because we have implemented a change in the habitat suitability map (i.e. an "environmental change event") every 5 years, you will notice that values in this column only change every 5th line. This makes sense, because under the unlimited dispersal scenario, all pixels that are suitable are colonized immediately when they become suitable. Thus, all suitable pixels become colonized during the first year (first dispersal event) when the habitat becomes suitable and there is nothing left to colonize during the remaining four dispersal events of an environmental change event.
NoDispersal	This column contains the number of pixels that would be occupied in the case of the "No Dispersal" scenario (i.e. a scenario in which we consider the dispersal of a species as null). The first line of this column contains the number of pixels of the species' initial distribution. Because we have implemented a change in the habitat suitability map (i.e. an "environmental change event") every 5 years, you will notice that values in this column only change every 5th line. This makes sense, because under the no dispersal scenario, all pixels are lost immediately when they turn unsuitable. Thus, all pixels that will be lost due to environmental change are lost during the first year (dispersal event) when the habitat becomes unsuitable and nothing more can be lost during the remaining four dispersal events of an environmental change event. *Note: by definition, under the "No Dispersal" scenario, a species can never increase in distribution. Thus the values in this column will only decrease, and never increase.*
Colonized	Number of pixels that are in a "Colonized" state at the end of the given dispersal event loop.
VegetativeResilient	Number of pixels that are in a "Vegetative Resilient" state at the end of the given dispersal event loop.
SeedBankResilient	Number of pixels that are in a "Seed Bank Resilient" state at the end of the given dispersal event loop.
Absent	Number of pixels that are in an "Unoccupied" state at the end of the given dispersal event loop.
LoopColonized	Number of pixels that turned into "Colonized" state during the given dispersal loop.
LoopVegetativeResilient	Number of pixels that turned into "Vegetative Resilient" state during the given dispersal loop.
LoopSeedBankResilient	Number of pixels that turned into "Seed Bank Resilient" state during the given dispersal loop.
LoopDecolonized	Number of pixels that turned into "Unoccupied" state during the given dispersal loop.

LostUnsuitaIble Number of pixels that turn unsuitable during the dispersal loop. Because in our simulation habitat suitability is updated every 5 years (5 dispersal steps), pixels can turn unsuitable only every 5 years and you will thus have many lines in your output file that are filled with "0" values.

Note that this values gives you the number of pixels that turn unsuitable, but they do not go extinct immediately as they first enter temporary resilience and then vegetative and seed-bank resilience (since we have chosen to implement these two options into our test simulation).

LostExtinction Number of pixels that have been lost during the dispersal loop due to a "pixel decolonization" event (specified through the P_{Ext} parameter).

LDDsuccess Number of successful LDD events (i.e. LDD events that led to the colonization of a pixel) that occurred during the dispersal loop.

VegResilRecovery Number of pixels that have recovered from a "Vegetative Resilient" state during the given dispersal loop (i.e. pixels that have changed their state from "Vegetative Resilient" to "Colonized").

SeedBankRecovery Number of pixels that have recovered from a "Seed Bank Resilient" state during the given dispersal loop (i.e. pixels that have changed their state from "Seed Bank Resilient" to "Colonized").

Note: Some type of pixel transitions (e.g. transition from "Colonized" to "Vegetative Resilient") or other events that are recorded in the "MigClim_FullOutput.txt" file can only occur at the time when the habitat suitability map is updated. Therefore, you will find that some lines of the "MigClim_FullOutput.txt" file that do not correspond to a environmental change event will necessarily have a value of "0".
In the particular case of the test example that we have been through in this user guide, we decided to update the habitat suitability every 5 years. You will therefore notice that for some columns (e.g. the "LoopVegetativeResilient" column) there is always one line with some value, followed by 4 lines with a value of "0". These four lines with a value of "0" correspond to dispersal events where no update in habitat suitability occurred.

At the end of the file, you will find a section entitled "MIGCLIM_SIMULATION_SUMMARY" and that contains, as its name suggests, a summary of your simulation. The values indicated are values for the entire simulation. The name of each value is fairly self-explanatory I think. The last line indicates the total time needed for the simulation to run. Here is an example of this section:

```
**********************************************************************************************************

MIGCLIM_SIMULATION_SUMMARY

Number_of_Initally_Colonized_Pixels:      43777

Number_of_Initally_Colonized_Pixels_that_Remained_Suitable(i.e. NoDispersal):       17574

Number_of_Pixels_Colonized_given_Unlimited_Dispersal(i.e. Universal_Dispersal):      61048

Number_of_Pixels_Colonized_at_End_of_Simulation:       47650

Number_of_Pixels_in_'Vegetative_Resilient'_stage_at_End_of_Simulation:            2326

Number_of_Pixels_in_'SeedBank_Resilient'_stage_at_end_of_Simulation:  0

**********************************************************************************************************

MigClim_Simulation_time:   00:01:08  hh:mm:ss

**********************************************************************************************************
```

MigClim_SummaryOutput.txt:

This file contains a summary of the simulation's output. It is particularly interesting if you have been running several simulations in a row, because it will then contain one line for each of the simulation that you have run, allowing you to have all your results nicely arranged in one single table.

To interpret this file, I would again advise to open it in Microsoft "Excel" or OpenOffice "Calc" using a space (" ") as delimiter. In this way, the data will be nicely displayed in different columns.

The following table will give you a detailed description of what the values of each column mean:

Column Name	Description
Simulation_Name	Name of the simulation.
Initial_Count	Number of pixel of the species' initial distribution
NoDispersal_Count	Number of pixels colonized at the end of the simulation under a "No Dispersal" scenario.
UnlimitedDispersal_Count	Number of pixels colonized at the end of the simulation under a "Unlimited Dispersal" scenario.
Colonized_Count	Number of pixels in "Colonized" state at the end of the simulation.
VegetativeResilient_Count	Number of pixels in "Vegetative Resilient" state at the end of the simulation.
SeedBankResilient_Count	Number of pixels in "Seed Bank Resilient" state at the end of the simulation.
Absent_Count	Number of pixels in "Unoccupied" state at the end of the simulation.
Total_Colonized	Total number of pixels colonized during the entire simulation.
Total_UnsuitLoss	Total number of pixels lost due to habitat turning unfavorable during the entire simulation.
Total_ExtLoss	Total number of pixels lost due to "Pixel Extinction" (i.e. specified through the P_{Ext} parameter) during the entire simulation.
Total_LDD	Total number of successful LDD events (i.e. LDD events that led to the colonization of a pixel) that occurred during the entire simulation.
Total_VegResilRecovery	Total number of pixels that have recovered from a vegetative resilience state during the entire simulation.
Total_SeedBankRecovery	Total number of pixels that have recovered from a seed bank resilience state during the entire simulation.
Simul_Time	Total time to run the simulation.

MigClim_Scenario_Log.txt:

This file contains all the information of your simulation and is also a file that can be used to launch your simulation again using the MIGCLIM "Multiple Scenario" mode (i.e. batch mode).

Note: if you launch your simulation again, make sure that you changed the output's name or deleted the output that you created before. This is because MIGCLIM will not over-write any raster output data, instead it will give you an error message saying that the output already exists (text files are not a problem as your new results will simply be appended to the existing file so you don't need to worry about loosing any results).

Note: you can use this file as a template to create your own "Multiple Scenario" file that contains all the simulations that you want to run at once.

Note: Any line of this file that starts with a # character will be ignored. This allows you to write some comments in the file, e.g. describing what kind of simulation the next line will perform.

Column Name	Description	Example from the test simulation
EnvChange_Steps	Number of "Environmental Change" events to be implemented in the simulation. Note that each "Environmental Change" event must correspond to one "Habitat suitability map". *Note: to run a simulation without environmental change, set this value equal to "0".*	10
Dispersal_Steps	Number of "Dispersal events" to be run within each "Environmental Change" event. *Note: In the case where the number of "Environmental Change" events was set to zero, the number of dispersal steps is equal to the total number of dispersal events that will occur during the simulation.*	5
InitialDist_Raster	Full path and name of the initial distribution of the species (ESRI grid file).	C:\MigClim\TestData\initial_dist
FirstHSmap_Raster	Full path and name of the first habitat suitability map of your simulation. If you have chosen to implement more than one environmental change step, then it is required that the first habitat suitability map has a name that ends with the number "1" (e.g. "hsmap1"). The subsequent habitat suitability maps are required to have the same name as the first one, with the number at the end indicating the order in which the should be considered (e.g. "hsmap2", "hsmap3", "hsmap4", etc...).	C:\MigClim\TestData\hsmap1
DispersalKernel_File	Full path and name of the "Dispersal Kernel" file.	C:\MigClim\TestData\DispersalKernel.txt
PixelSizeFactor	Value of the "Pixel Size Factor". *Note: If you do not know what this parameter is for, set its value to "1" (this is the default value).*	1
MinLDD_Dist	Minimum LDD distance in pixel units.	6
MaxLDD_Dist	Maximum LDD distance in pixel units.	40
LDD_Freq	Frequency of LDD events *(P_{LDD})*.	0.01
ReprodKernel_File	Full path and name of the "Reproductive Potential Kernel" file.	C:\MigClim\TestData\ReproductivePotentialKernel.txt

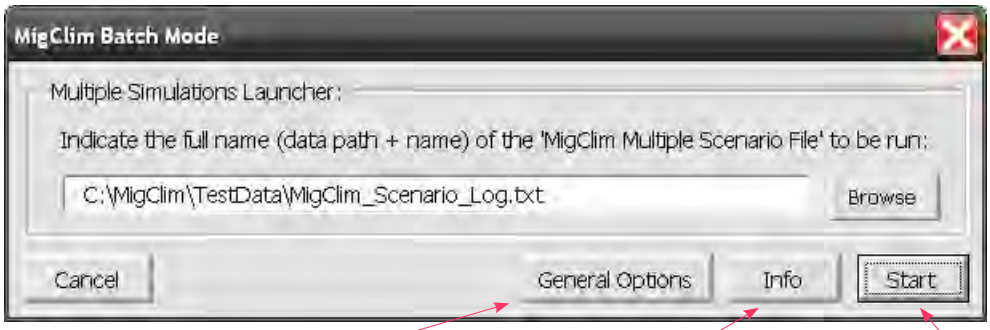
Resilience_File	Full path and name of the "Resilience Kernel" file.	C:\MigClim\TestData\ResilienceKernel.txt
Barrier_Raster	Full path and name of the "Barrier" file (ESRI grid format). *Note: This parameter is optional. To run a simulation without a "Barrier" feature, set this parameter to "NA".*	C:\MigClim\TestData\test_barrier
Barrier_Type	Type of barrier to be used in the simulation. This parameter can have only three values: "strong", "weak" or "NA" if no barrier feature is used in the simulation.	strong
Filter_Raster	Full path and name of the "Filter" grid (ESRI grid format). *Note: This parameter is optional. To run a simulation without a "Filter" feature, set this parameter to "NA".*	C:\MigClim\TestData\test_filter
Extinction_Freq	Frequency of the "Pixel Extinction" parameter (P_{ext}).	NA
RiverAndRoad_File	Full path and name of the "River and Road file". *Note: This parameter is optional. To run a simulation without a "River and Road" dispersal, set this parameter to "NA".*	NA
DispersalDist_Rivers	Multiplication factor for dispersal distance along river features. *Note: If you do not wish to implement "river and road dispersal", set this parameter to "0".*	0
RiverLDD_Freq	Frequency of LDD events along river features. *Note: If you do not wish to implement "river and road dispersal", set this parameter to "0".*	0
RiverLDD_MinDist	Minimum LDD distance along rived features, in pixel units. *Note: If you do not wish to "implement river and road dispersal", set this parameter to "NA".*	NA
RiverLDD_MaxDist	Maximum LDD distance along rived features, in pixel units. *Note: If you do not wish to implement "river and road dispersal", set this parameter to "NA".*	NA
DispersalDist_Roads	Multiplication factor for dispersal distance along road features. *Note: If you do not wish to implement "river and road dispersal", set this parameter to "0".*	0
RoadLDD_Freq	Frequency of LDD events along road features. *Note: If you do not wish to implement "river and road dispersal", set this parameter to "0".*	0
RoadLDD_MinDist	Minimum LDD distance along road features, in pixel units. *Note: If you do not wish to implement "river and road dispersal", set this parameter to "NA".*	NA
RoadLDD_MaxDist	Maximum LDD distance along road features, in pixel units. *Note: If you do not wish to implement "river and road dispersal", set this parameter to "NA".*	NA
DEM_Raster	Digital Elevation Model raster that is needed if you wish to implement the slope weighted dispersal option. *Note: If you do not wish to implement "slope weighted dispersal", set this parameter to "NA".*	NA

SlopeFactor_File	Text file indicating how slope weights the dispersal distance. Only needed if you wish to implement the slope weighted dispersal option. *Note: If you do not wish to implement "slope weighted dispersal", set this parameter to "NA".*	NA
WindDirection_Raster	Raster dataset that gives the direction of wind for each pixel. Only needed if you wish to implement wind weighted dispersal. *Note: If you do not wish to implement "wind weighted dispersal", set this parameter to "NA".*	NA
WindSpeed_Raster	Raster dataset that gives the speed of wind for each pixel. Only needed if you wish to implement wind weighted dispersal. *Note: If you do not wish to implement "wind weighted dispersal", set this parameter to "NA".*	NA
MinWindSpeed	Minimum wind speed under which the wind does not affect dispersal. Only needed if you wish to implement wind weighted dispersal. *Note: If you do not wish to implement "wind weighted dispersal", set this parameter to "NA".*	NA
MaxWindSpeed	Maximum wind speed above which higher wind speed does not affect dispersal more than it already does. Only needed if you wish to implement wind weighted dispersal. *Note: If you do not wish to implement "wind weighted dispersal", set this parameter to "NA".*	NA
MinWindFactor	Dispersal distance multiplication factor associated with minimum wind speed. Only needed if you wish to implement wind weighted dispersal. *Note: If you do not wish to implement "wind weighted dispersal", set this parameter to "NA".*	NA
MaxWindFactor	Dispersal distance multiplication factor associated with maximum wind speed. Only needed if you wish to implement wind weighted dispersal. *Note: If you do not wish to implement "wind weighted dispersal", set this parameter to "NA".*	NA
Wind_Freq	Frequency with which the wind is affecting dispersal. Only needed if you wish to implement wind weighted dispersal. *Note: If you do not wish to implement "wind weighted dispersal", set this parameter to "NA".*	NA
Output_Folder	Output folder where all MigClim outputs will be saved.	C:\MigClim\TestData
Output_Name	Name of the simulation.	simul1
Simulation_Repeat	Number of times a simulation should be repeated/replicated.	1
SaveRasterEvery	Frequency with which the MigClim output should be saved to a map (i.e. an ESRI grid). Note that in all cases all MigClim outputs will always be saved in the form of a text file, but if you want to see the spatially explicit result of your simulation, it needs to be saved as a map. If you set this parameter to "1", then a map of your output will be saved for each repetition/replicate of your simulation. If you set the parameter equal "2", a map will be saved for every second repetition/replicate of your simulation. If you set the parameter equal to "3", a map will be saved every third repetition/replicate of your simulation. And so on.	1

Part 4: Additional options available in MigClim

Multiple Simulation Mode (MigClim batch mode)

Once you will get familiar with MigClim and its various parameters, chances are that you will quickly become tiered of using the "MigClim Wizard" that we went through in our test example (see first part of this user guide). This is the time when you will really appreciate the "Multiple Simulation Mode", that allows you to launch a virtually infinite number of simulations with just a few mouse clicks.



Use this button to access the "General Options" of MigClim. These are the same options than those you access through the "Options" > "General Options" menu.

This button displays information about the simulations that are contained within the Multiple Scenario File that you selected.

Click here to launch you simulations.

Launching your simulations in MigClim's batch mode could hardly be any easier: simply browse to your "MigClim Multiple Scenario File" using the ***Browse*** button, or copy/paste the full file path and name. All you have to do next is hit the ***Start*** button. MigClim will first check whether all your simulations are correct (i.e. the data they refer to is there and the parameters have been given adequate values). Depending on how many simulations you have in your "MigClim Multiple Scenario File" this might take a little while.

Once all simulations contained in your "MigClim Multiple Scenario File" have been proofed, a dialog box will appear, telling you whether a problem has been detected or not. With a bit of luck no problems are detected, and you can simply click the ***OK*** button on the dialog box and your simulations will be launched.

Depending on how many simulations you have asked for, you will have time to go and have a coffee, two coffees, go for lunch, take a day off, a month of holidays or an entire sabbatical year! So make sure you run some tests to find out more or less how long your computing will take.

If an error was detected in your "MigClim Multiple Scenario File", then the simulations will obviously not be started. Instead, MigClim will produce a file called "ErrorLog.txt" in your output folder. This text file contains (more or less intelligible) indications about what is wrong with your "MigClim Multiple Scenario" File" and should hopefully provide you with some guidance to fix your file.

Coming soon... stay tunned !

Road and River dispersal

Weighting dispersal with wind direction and strength

Weighting dispersal with terrain slope

Pixel Size Factor