UNIL | Université de Lausanne
Faculté de biologie et de médecine
Département d'écologie et d'évolution

Predicting spatial patterns of plant biodiversity: from species to communities

Thèse de doctorat ès sciences de la vie (PhD)

Présentée à la
Faculté de biologie et médecine de l'Université de Lausanne
Par

Anne Dubuis

Master en biologie de l'Université de Lausanne

Jury

Prof. Stefan Kunz, Président
Prof. Antoine Guisan, Directeur de thèse
Dr. Pascal Vittoz, Expert
Prof. Jean-Claude Gégout, Expert
Prof. Miska Luoto, Expert

Lausanne 2013

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. Stefan **Kunz**
Directeur de thèse	Monsieur	Prof. Antoine **Guisan**
Experts	Monsieur	Dr Pascal **Vittoz**
	Monsieur	Prof. Jean-Claude **Gégout**
	Monsieur	Prof. Miska **Luoto**

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

Madame Anne Dubuis

master en biologie de l' Université de Lausanne

intitulée

Predicting spatial patterns of plant biodiversity: from species to communities

Lausanne, le 8 mars 2013

pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Stefan Kunz

SUMMARY

Understanding the distribution and composition of species assemblages and being able to predict them in space and time are highly important tasks to investigate the fate of biodiversity in the current global changes context. Species distribution models are tools that have proven useful to predict the potential distribution of species by relating their occurrences to environmental variables. Species assemblages can then be predicted by combining the prediction of individual species models.

In the first part of my thesis, I tested the importance of new environmental predictors to improve species distribution prediction. I showed that edaphic variables, above all soil pH and nitrogen content could be important in species distribution models. In a second chapter, I tested the influence of different resolution of predictors on the predictive ability of species distribution models. I showed that fine resolution predictors could ameliorate the models for some species by giving a better estimation of the micro-topographic condition that species tolerate, but that fine resolution predictors for climatic factors still need to be ameliorated.

The second goal of my thesis was to test the ability of empirical models to predict species assemblages' characteristics such as species richness or functional attributes. I showed that species richness could be modelled efficiently and that the resulting prediction gave a more realistic estimate of the number of species than when obtaining it by stacking outputs of single species distribution models. Regarding the prediction of functional characteristics (plant height, leaf surface, seed mass) of plant assemblages, mean and extreme values of functional traits were better predictable than indices reflecting the diversity of traits in the community. This approach proved interesting to understand which environmental conditions influence particular aspects of the vegetation functioning. It could also be useful to predict climate change impacts on the vegetation.

In the last part of my thesis, I studied the capacity of stacked species distribution models to predict the plant assemblages. I showed that this method tended to over-predict the number of species and that the composition of the community was not predicted exactly either. Finally, I combined the results of macro-ecological models obtained in the preceding chapters with stacked species distribution models and showed that this approach reduced significantly the number of species predicted and that the prediction of the composition is also ameliorated in some cases. These results showed that this method is promising. It needs now to be tested on further data sets.

RESUMÉ

Comprendre la manière dont les plantes se répartissent dans l'environnement et s'organisent en communauté est une question primordiale dans le contexte actuel de changements globaux. Cette connaissance peut nous aider à sauvegarder la diversité des espèces et les écosystèmes. Des méthodes statistiques nous permettent de prédire la distribution des espèces de plantes dans l'espace géographique et dans le temps. Ces modèles de distribution d'espèces, relient les occurrences d'une espèce avec des variables environnementales pour décrire sa distribution potentielle. Cette méthode a fait ses preuves pour ce qui est de la prédiction d'espèces individuelles. Plus récemment plusieurs tentatives de cumul de modèles d'espèces individuelles ont été réalisées afin de prédire la composition des communautés végétales.

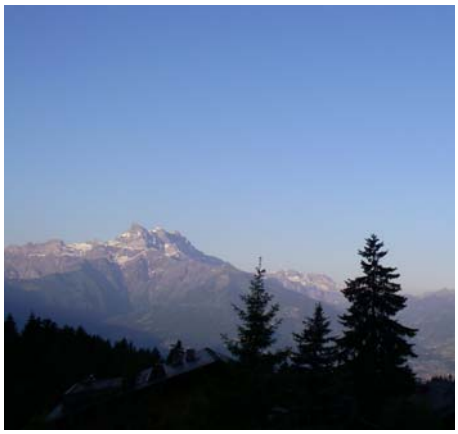
Le premier objectif de mon travail est d'améliorer les modèles de distribution en testant l'importance de nouvelles variables prédictives. Parmi différentes variables édaphiques, le pH et la teneur en azote du sol se sont avérés des facteurs non négligeables pour prédire la distribution des plantes. Je démontre aussi dans un second chapitre que les prédicteurs environnementaux à fine résolution permettent de refléter les conditions micro-topographiques subies par les plantes mais qu'ils doivent encore être améliorés avant de pouvoir être employés de manière efficace dans les modèles.

Le deuxième objectif de ce travail consistait à étudier le développement de modèles prédictifs pour des attributs des communautés végétales tels que, par exemple, la richesse en espèces rencontrée à chaque point. Je démontre qu'il est possible de prédire par ce biais des valeurs de richesse spécifiques plus réalistes qu'en sommant les prédictions obtenues précédemment pour des espèces individuelles. J'ai également prédit dans l'espace et dans le temps des caractéristiques de la végétation telles que sa hauteur moyenne, minimale et maximale. Cette approche peut être utile pour comprendre quels facteurs environnementaux promeuvent différents types de végétation ainsi que pour évaluer les changements à attendre au niveau de la végétation dans le futur sous différents régimes de changements climatiques.

Dans une troisième partie de ma thèse, j'ai exploré la possibilité de prédire les assemblages de plantes premièrement en cumulant les prédictions obtenues à partir de modèles individuels pour chaque espèce. Cette méthode a le défaut de prédire trop d'espèces par rapport à ce qui est observé en réalité. J'ai finalement employé le modèle de richesse en espèce développé précédemment pour contraindre les résultats du modèle d'assemblage de plantes. Cela a permis l'amélioration des modèles en réduisant la sur-prédiction et en améliorant la prédiction de la composition en espèces. Cette méthode semble prometteuse mais de nouveaux tests sont nécessaires pour bien évaluer ses capacités.

TABLE OF CONTENT

INTRODUCTION



THE STUDY OF SPECIES DISTRIBUTION

Understanding how species are distributed and how they assemble to form communities and ecosystems is a question that has motivated scientific interest for a long time. The on-going global changes – climate warming, destruction of habitats, pollution and biological invasions - are impacting strongly on biodiversity (Sala et al., 2000; Pereira et al., 2010; Chen et al., 2011; Cardinale et al., 2012). These changes increase drastically the need of knowledge that could be used to anticipate and prevent deleterious effects of global changes on biodiversity and ecosystem functioning. Regarding plants, phytogeography a discipline merging botany and geography, has investigated the distributions of plant species or communities and their respective drivers (climatic, edaphic, biotic) since the beginning of the 19th century. In 1805, Alexander von Humbolt and Aimé Jacques Bonpland, two botanists that were just returning from a trip through South America, published a book entiltled “Essai sur la géographie des plantes”. In this essay, they considered several environmental factors like temperature, humidity and geology to explain the repartition of species, thus setting early bases of modern biogeography. In 1820 Augustin Pyramus de Candolle, followed by his son Alphonse defined botanical regions based on vegetation form and their relations to climatic region (Magnin-Gonze, 2004). These precursors were followed by numerous scientists that studied relationships between living organisms and their environment.

Mathematical models are increasingly used in ecology, among others to understand and predict relations between species and their environment. Models have divergent qualities and can be separated in three groups that each maximise two out of the three following properties: generality, reality and precision (Levins, 1966; Guisan & Zimmermann, 2000). Analytical or mathematical models favour generality and precision. They allow predicting a response within a simplified reality. Lokta-Volterra equation is well known example of analytical model (Volterra, 1926). The second group, mechanistic models (also called process-based or physiological models) are realistic and general. They give predictions based on cause-effect relationships that are considered as biologically functional. Examples are numerous and include gap models used in forest ecology (Bugmann, 2001), metapopulation models (Akçakaya et al., 2004) or trait-based community assembly models (Shipley et al., 2006). The third group, empirical models are considered as realistic and precise. Species distribution models are one kind of empirical model. They use statistical techniques to relate

species occurrences to environmental variables and predict their distribution (Guisan & Zimmermann, 2000). They are static and probabilistic opposed to the mechanistic models mentioned above. They allow performing large-scale predictions while being far less data demanding than other sort of models. These models bear different names throughout the literature such as species distribution models, ecological niche models, niche-based models, predictive habitat distribution models or bioclimatic envelope models. I will use the appellation “species distribution model” (SDM) in this thesis as this is one of the most current.

SPECIES DISTRIBUTION MODELS

The niche concept

Species distribution models rely on the niche concept, first defined by Grinnell (1917) as all the sites where organisms of a species can live, and later refined by Hutchinson (1957) who defined the fundamental (physiological) niche as a *n*-dimensional environmental volume where conditions allow positive growth of populations. Hutchinson further distinguished the realized niche observed *in situ* as a sub-space of the fundamental niche where the species is not excluded by negative biotic interactions. These interactions are mainly competition in Hutchinson writings, but modern definitions include also predation, pathogens, or absence of symbionts (Soberón, 2007). Consequently the realised niche should, by definition, be narrower than the fundamental niche (but see Araújo & Guisan (2006) for the potential role of positive biotic interactions), and empirical data can only inform on the realized niche. Therefore, SDMs that relate observed species’ presences to environmental conditions do estimate the realized environmental niche of species (Guisan & Thuiller 2005, Araujo & Guisan 2006).

Assumptions in species distribution modelling

Using SDMs requires making methodological and theoretical assumptions (Araújo & Guisan, 2006; Wiens et al., 2009), some of which can limit the application of models in some cases (see Pulliam 2000; Jiménez-Valverde et al. 2008 for reviews). Here, I detail two of them that prove important for understanding some of my results.

First of all, due to the fact that data used to build such a model has been collected during a limited period of time (snapshot), we assume that the modelled species is in equilibrium with its environment during this period. Indeed, a species that occupies all favourable places and which distribution will not be modified in the future will allow to

build a model that describes its niche in the best possible way. It is known that this assumption is not fulfilled when the modelled species is still colonizing new habitats, for example in the case of post glacial recolonisation (e.g., Leathwick 1998) or biological invasions (e.g. Petitpierre et al. 2012). Environmental stochasticity can also be a cause of disequilibrium between a species distribution and its environment (Pulliam, 2000), but it is unlikely to break the overall equilibrium across large areas.

A second problem when using SDMs is to know what type of niche is modelled (Araújo & Guisan, 2006). As occurrences come most often from field sampling, it is assumed that the realized niche is modelled, as the observed distribution on the modelled species is already constrained by biotic interactions. However, it is not possible from a simple field observation to know if a species will be able to continue to grow at the place where it has been observed or whether it is part of a sink population (Pulliam, 2000), for example. Moreover as no field sampling can be exhaustive, we do not know if all biotic interactions are taken into account. Consequently we assume that what SDMs predict is an estimate of the realised niche and that it cannot be considered as a true description of a species niche (Soberón, 2007).

Data to build SDMs

The basis of SDM building is species data. These data can be presences only, presences - absences, or abundances data and modelling techniques exist for each of this data type. Presences-absences data are used if exhaustive inventories are available and mostly in the case of sessile organisms for which the observed absences can be considered as reliable. If reliable absences are not available (study of a mobile species or data obtained from multiple sources), models can be built using pseudo-absences (e.g., Barbet-Massin et al. 2012). Abundances data are not widely used but they have been giving promising results (Randin et al., 2006; Van Couwenberghe et al., 2013) allowing more sharpness in the habitat suitability estimation.

Predictor data, used as explanatory variable in the model, should be as proximal as possible, for the model to have a real ecological signification (Austin, 2002, 2007). They must have a direct impact (ecological or physiological) on species or represent a resource necessary for its growth. Predictors are often climatic and/or topographic as these are relatively easy to obtain, but can also include land-use (Randin et al., 2009b), geomorphology (Randin et al., 2009c) or biotic interaction through the distribution of a competing species for example (Leathwick & Austin, 2001; Pellissier et al., 2010a; Le Roux et al., 2013). While the quality and accuracy of predictors is important to obtain

ecologically relevant models, such data are not always available in a spatially explicit form, thus making the model projection difficult. Edaphic data among other, are known to influence species distribution, but are rarely integrated in SDM due to their low availability. Another question regarding predictors is the resolution (or grain size) to use for model calibration and projection. Ideally the resolution of predictors should be the same as the resolution of the data to model (Guisan & Thuiller, 2005). But this is not always possible. The resolution should at least be in accordance with the study aims and the process that is studied. For example, at a too coarse resolution, the results of models can lead to erroneous conclusions because particular phenomenon are overlooked (Randin et al., 2009a). This is especially true for sessile organisms such as plants (Guisan & Thuiller, 2005).

Model calibration

The relation between species data and predictors is then inferred with a statistical modelling technique. Different algorithms can be used in this context (Franklin, 2009), such as regression approaches (e.g. generalised linear and generalized additive models (McCullagh & Nelder, 1989; Hastie & Tibshirani, 1990), classification trees (e.g. boosted regression trees, Ridgeway 1999; Friedman 2001; or random forest, Breiman 2001), environmental envelopes (Busby, 1991), ecological niche factor analysis (Hirzel et al., 2002), artificial neural networks (Ripley, 1996), or maximum entropy (Phillips et al., 2006). The performance of these different techniques is compared in Elith et al. (2006). The results obtained for a similar species with different algorithms can be different (Elith et al., 2006; Grenouillet et al., 2011), consequently the algorithm used should be chosen according to the aims of the study. Different algorithms should be used to produce a model that explains the data well or that can be projected reliably in another context (Randin et al., 2006). A widely recommended solution is to use an ensemble forecasting approach to derive predictions from the models (Araújo et al., 2007). During ensemble forecasting, the projections resulting from the different algorithms are averaged, possibly weighted by their respective evaluation results. This allows obtaining more robust predictions and estimating the variability across techniques (Buisson et al., 2010).

Model evaluation and projection

After calibration, the quality of the model should be evaluated. Several metrics exist that quantify the difference between predicted and observed data. Among currently used metrics we can cite the area under the curve of a receiver operating characteristic

curve (AUC; Hanley & Mcneil 1982; Fielding & Bell 1997), Cohen's Kappa (Cohen, 1960) and the true skill statistics (TSS, Allouche et al. 2006). In theory, those metrics should be computed on an independent dataset. As independent data are rarely available, it is current to use one part of the original dataset left aside from model calibration to evaluate the model. For example, during a repeated split sample procedure, as implemented in BIOMOD (Thuiller et al., 2009), the calibration dataset is split in two, the larger part (e.g. 70%) is used for calibration of the model that will be then evaluated on the remaining part (e.g. 30%) of the data. This procedure is repeated and the evaluation results are averaged.

Finally, models can be projected, either in the same area and period or in different space and time, e.g., to build projection of the species after climate warming. The output of a model projection is expressed as the probability of the species to occur in the pixels considered. For numerous SDM applications, the transformation of probabilities of presence to potential presence and absences is needed. This can be achieved for example by maximizing threshold based evaluation measures as Kappa or TSS, or by maximizing the sensitivity (presences predicted as such) and specificity (absences predicted as such) of the model or as to reflect the prevalence of presences and absences of the calibration dataset. See Liu et al. (2005) for a review and comparison of those methods

Use of species distribution models

SDMs have been applied to a great variety of organisms: from viruses (Machado-Machado, 2012) and phytoplankton (Hallstan et al., 2012) to vascular plants (Meier et al., 2010; Pellissier et al., 2010a; Engler et al., 2011) and lichens (Bergamini et al., 2007) to insects (Maggini et al., 2002; Lütolf et al., 2006; Marmion et al., 2009; Titeux et al., 2009), birds (Wisz et al., 2007), fishes (Sundblad et al., 2009; Jones et al., 2012; Martin et al., 2012) and mammals (Boitani et al., 2007; Vega et al., 2010; Rondinini et al., 2011). In addition to quantifying the environmental niche of species (Guisan et al., 1998; Coudun & Gegout, 2007; Rondinini et al., 2011), SDMs have been used to test ecological (Pellissier et al., 2010a; Petitpierre et al., 2012) or evolutionary hypotheses (Vega et al., 2010; Schorr et al., 2012), to assess the impact of future climate changes (Engler et al., 2011) and/or land use changes (Dirnböck et al., 2003; Pompe et al., 2008; Vicente et al., 2011) on species distribution. SDM can also be used to project distribution of species in the past (hindcasting; Pearman et al. 2008; Maiorano et al. 2012). In more applied perspectives, SDMs are used to support management plans for endangered species (Cianfrani et al., 2011), help to find new occurrences of rare

species (Engler et al., 2004; Guisan et al., 2006; Le Lay et al., 2010), support conservation planning and reserve selection (Carroll, 2010; Carvalho et al., 2010; Tognelli et al., 2010) or predict the proliferation of invasive species (Thuiller et al., 2005). They have also been applied in an economic perspective to predict the distribution of commercial fish species (Jones et al., 2012), suitable habitat for baobab tree cultivation (Sanchez et al., 2010) or the fate of coffee farming under climate change (Davis et al., 2012). SDMs use mostly species occurrences as response variable. But other response variables can be predicted with similar techniques, for example vegetation types (Maggini et al., 2006; Baselga & Araújo, 2009), species richness (Wohlgemuth, 1998; Pineda & Lobo, 2009) or functional characteristics of species assemblages (Pellissier et al., 2010c; Küster et al., 2011). SDMs can also be summed to predict species assemblages from individual predictions; all these cases will be developed in the following part of this introduction.

THE COEXISTENCE OF SPECIES

What is a community?

Species do not occur alone, and when looking around us, we can observe assemblages of different species. For a long time, there has been a debate in vegetation ecology on how to comprehend a community. It started between F. E. Clements and H. A. Gleason, two ecologists with completely different opinion on the subject. Clements saw the community as an entity, a “super-organism”, where species were strongly interacting and that could persist through time (Clements, 1916, 1936). Gleason challenged this view with a more dynamic approach, arguing that a community was only the sum of organisms sharing a similar environment at a particular time (Gleason, 1926). These two mind-sets evolved in parallel to generate, on the one hand, community ecology and on the other hand biogeography (Hortal et al., 2012). These two discipline have rarely been merged, and the debate about the existence and assembly of communities is still on-going (Ricklefs, 2008; Brooker et al., 2009). It extended on the existence of rules allowing to explain community composition (Weiher & Keddy, 1999).

Assembly rules?

The concept of assembly rules states that species do not co-occur randomly and that, negative and positive interactions can be detected between species growing in a similar environment. Despite numerous studies looking for assembly rules their existence is still debated (Grime, 2006; Wilson, 2007). Numerous methods have been used to detect assembly rules (see Götzenberger et al. 2011 for a review, Annex 4). These methods all have in common the comparison of observed patterns with patterns obtained after randomizing the data with a null model (Gotelli, 2000). The measures used to detect assembly rules can be based on species. Co-occurrences (Dullinger et al., 2007; Maestre et al., 2008), for example, are measured with indices that detect spatial segregation or aggregation in a species by site matrix. Niche limitation (Wilson et al., 1987) is evaluated by comparing the observed variation in species richness to what can be expected at random, given that the number of species should be limited by the number of available niches. Functional traits can also be used to detect assembly rules. Limiting similarity (MacArthur & Levins, 1967; Wilson & Stubbs, 2012) investigates the differences in term of functional traits between species of a community by testing if the observed trait variation is different from what is expected by chance. Traits divergence is interpreted as evidence for limiting similarity: species that are

different in their traits can coexist more easily as they compete less with each other; while trait convergence reflect environmental filtering: species share functional traits that give them particular environmental tolerance (Cornwell et al., 2006; Cornwell & Ackerly, 2009). Such traits patterns must be considered cautiously as they depend on the scale at which they are considered and on the traits under study. A convergence of traits can also be due to competitive exclusion of less competitive species (Mayfield & Levine, 2010; de Bello et al., 2012). See the paragraph “Functional traits as species substitutes” for more details on functional traits.

Niche-based or neutral community assembly?

More recently a new controversy has opposed the niche-based view of community assembly with the neutral theory of community. The niche-based theory, derived from Hutchinson’s niche definition (1957), states that communities are made up from species that have environmental requirement allowing them to occur in a similar place. Each species occupy only the sites for which it is best adapted and from which it is able to exclude competitors (Gause, 1936). On the other hand, neutral theory states that most species are able to grow at most sites (they are equivalent in term of fitness) and that chance, dispersal, extinction and speciation are the major process of community formation (Hubbell, 2001, 2005; Chave, 2004). Neutral models have been able to reproduce patterns such as species area relationships and patterns of relative abundance across species (Hubbell, 2001), but not assemblage composition. Currently, the prevailing view argue for a reunification of both theories (Leibold & McPeek, 2006; Adler et al., 2007; Wennekes et al., 2012) as the process they cover can be seen as complementary. The niche assembly perspective is a small-scale but precise approach; it relies on processes occurring at the individual level. The neutral perspective relies on more broad scale processes (dispersal, chance) and highlight phenomenon that are often neglected in classical uses of niche-based theory (Wennekes et al., 2012). Frameworks aiming at merging both theories have recently been proposed (Stokes & Archer, 2010; Tang & Zhou, 2012).

Functional traits as species substitutes

Rather than being defined as one taxonomical entity, a species can also be summarized through a sum of functional traits, Functional traits are defined as any morphological, physiological or phenological feature that is measurable at the individual level and affects individual performances; a trait can be understood as a surrogate of a function of the species (Violle et al., 2007). The traits of a plant relate with its

environments in two ways (Lavorel & Garnier, 2002; Lavorel et al., 2007). They can either be seen as the organism's response to a particular set of environmental conditions (i.e., response trait), or they can be seen as having an effect on ecosystem functioning (i.e., effect trait). Functional traits have been widely used in ecology. They allow classifying species in functional groups reacting to or affecting the environment in a similar way (Lavorel & Garnier, 2002). Functional groups can be used as response variables in empirical (Steinmann et al., 2009), or dynamic models (Bonan et al., 2003; Sitch et al., 2003). If traits are summarized at the community level through indices like community weighted means or functional diversity (Mouchet et al., 2010; Schleuter et al., 2010), they can be used to functionally characterize plant communities in different environments (Sonnier et al., 2010), or to understand community assembly by identifying convergence-divergence assembly rules (Cornwell et al., 2006; Cornwell & Ackerly, 2009; de Bello et al., 2009; Venn et al., 2011). They can also be used as constraints to predict communities composition (Shipley et al., 2006; Douma et al., 2012). The same functional indices have been related to ecosystem functions such as net above-ground productivity, pasture grass quality and litter decomposition (Hector et al., 1999; Garnier et al., 2004; Spehn et al., 2005; Pontes et al., 2007; Mokany et al., 2008; Fortunel et al., 2009).

An important advantage of using functional traits is that they provide greater generality and predictability in understanding the creation and functioning of ecosystems than approaches based on species identity alone (Diaz & Cabido, 2001; Hooper et al., 2005). They can reflect the effects of the environment on species but also the effect of dominant species or the redundancy of species with similar functional traits on ecosystem functioning.

SPECIES ASSEMBLAGE MODELS

Although the status of communities is not clear and still debated, numerous attempts have been made to build models of community from assembling individual species. To model communities, the data needed are basically the same that are used to model single species distributions: spatially explicit presences and absences of the species forming the community as well as environmental predictors. Here, a crucial requirement is that all species must have been inventoried within each plot during the same survey.

In their 2006 paper, Ferrier and Guisan review three different strategies (see below) that can be used to model biodiversity at the community level. These strategies process the data differently, provide different outputs and have diverse strengths and weaknesses. The choice of one technique over the other should depend on the type of dataset to analyse and the aims of the study.

Assemble first, predict later

The first strategy is called “Assemble first, predict later” and corresponds to a Clements view of the communities, considered as entities. The first stage of this strategy consists in assembling the species data by classification, ordination or aggregation. This is done on the species data only, without reference to environment. In a second step, the entities generated previously are modelled as a function of the environment. Species groups originating from classification of ordination analysis (Steinmann et al., 2009), and community types (Maggini et al., 2006), can be predicted with these methods. Community characteristics can also be derived from community data and modelled. A common example is species richness that has been modelled in a great variety of studies and contexts (Wohlgemuth, 1998; Gould, 2000; Luoto et al., 2004). Some studies have also attempted to predict functional values at the community level (Pellissier et al., 2010c; Küster et al., 2011). This approach is interesting as it is quite efficient to analyse large datasets, it also allows rare species to be taken into account in the analysis. As drawbacks it does not allow individualistic species responses and does not provide individual species distributions, i.e. no predicted species list is available after modelling. The transferability of such model under different condition is also limited.

Predict first, assemble later

The second strategy is called “Predict first, assemble later” and matches more with Gleason’s view. Here the communities result from the overlap of independent species

distributions. The first step in this case is to model each species independently as a function of environmental predictors, to obtain one prediction map per species. These maps are then stacked to predict the species occurring together at each point. The resulting predicted communities can then be treated in the second stage as the observed community data in the first stage of the "Assemble first, predict later" strategy, i.e., submitted to classification, ordination or aggregation to derive the same kind of communities characteristics as mentioned above (species groups, community type or species richness for example). This strategy has been used to derive community composition (Peppler-Lisbach et al., 2004; Aranda & Lobo, 2011; Pottier et al., 2013), species richness (Guisan & Theurillat, 2000; Parviainen et al., 2009; Pineda & Lobo, 2009), phylogenetic diversity measures (Pio et al., 2011). This strategy is the most time consuming, as it requires calibrating one model per species. Another drawback is that rare species (with insufficient occurrence number) cannot be modelled (but see Lomba et al. 2010); communities predicted with this strategy are thus likely to be incomplete. However this strategy is interesting because it allows each species to respond differently to the environment, it provides one distribution map per species, i.e., one species list per plot. Another interesting quality of this approach is that it can be used to analyse data originating from different surveys, as each models are independent.

Assemble and predict together

The third strategy is called "Assemble and predict together". In this analysis species are assembled and modelled at the same time. This can be achieved thanks to constrained ordination techniques, which are able to group species and relate them to environmental predictors at the same time. This approach has been widely use in vegetation analysis (Virtanen et al., 2006; Pinto et al., 2007), but more rarely to predict species (Guisan et al., 1999) or patterns of biodiversity outside of the surveyed locations (Baselga & Araújo, 2009, 2010). Other techniques mainly used to model single species distributions also allow fitting "multiresponse" models to data for multiple species, e.g., vector generalized additive models (Yee & Mackenzie, 2002). This third strategy is particularly efficient when it comes to analysing important datasets. It also allows rare species to be taken into account in the analysis as they are considered together with the others. It addresses interactions between species but also allows individualistic species responses and part of the methods provides individual species distribution predictions. However the predictive power of some of the modelling

technique used to assemble and predict together have been questioned in two recent studies (Baselga & Araújo, 2009, 2010).

Spatially explicit species assemblage modelling: the SESAM framework

The three approaches presented above have complementary strength and weaknesses. The more important qualities are on one hand the ability to use as complete species data as possible, including rare species. On the other hand it is interesting also to obtain a prediction of the community composition (a species list) as a result, particularly if the modelling is done with conservation aims.

In this thesis, I will focus on the two first approaches. The approach “assemble first, predict later” called hereafter macroecological modelling (MEM), has been shown to produce realistic estimation of species richness (Algar et al., 2009; Newbold et al., 2009) but does not provide a species list as output. On the contrary, the second approach, “predict first assemble later”, hereafter called stacked species distribution models (S-SDM) has consistently been shown to overpredict the number of species (Algar et al., 2009; Newbold et al., 2009; Pineda & Lobo, 2009; Aranda & Lobo, 2011). On the basis of this idea, Guisan and Rahbek (2012) proposed to use macroecological models to limit the number of species predicted by S-SDM, thus bringing together the strength of both approaches. The hypotheses underlining both approaches are different. MEM implies that the environment determines the number of species that a unit of community can hold (or any other assemblage property). Communities are saturated and this saturation is defined by the environmental carrying capacity. On the other hand the S-SDM understands a community as an assemblage of all species sharing similar requirements. In this approach, species respond individually to the environment and the assumptions are the same than for SDMs. The predictors used for the two different types of models can be the same, but their meaning will be slightly different. For example, the temperature means the available energy for a community in MEM, when it defines the limit of a species range in SDM.

SESAM could be applied in four steps. First a species pool is defined, accounting for dispersal limitations. Second, the distribution of each chosen species is modelled individually and combined to construct stacked species distribution models. This corresponds to selecting the species according to abiotic conditions. Thirdly, an assemblage property (e.g., species richness) is predicted for the same area and the same species data-set with a MEM. It will be used to set a limit to the S-SDM if too much species are predicted. Finally, species entering the final assemblage are chosen

thanks to ecological assembly rules, if such rules have been developed for the species under study.

This framework can be seen as a solution to reconcile the view of Clements and Gleason to predict community composition (Hortal et al., 2012). The abiotic species pool is formed as a sum of species distribution prediction (Gleason view) at larger scale, while the choice of species being part of the assemblage in the end is based on biotic interaction or habitat selection at a more local scale, closer to Clements view.

The SESAM framework has not been tested yet, and several question still need to be answered before that. SDM and MEM should be developed, and their combination must be tested.

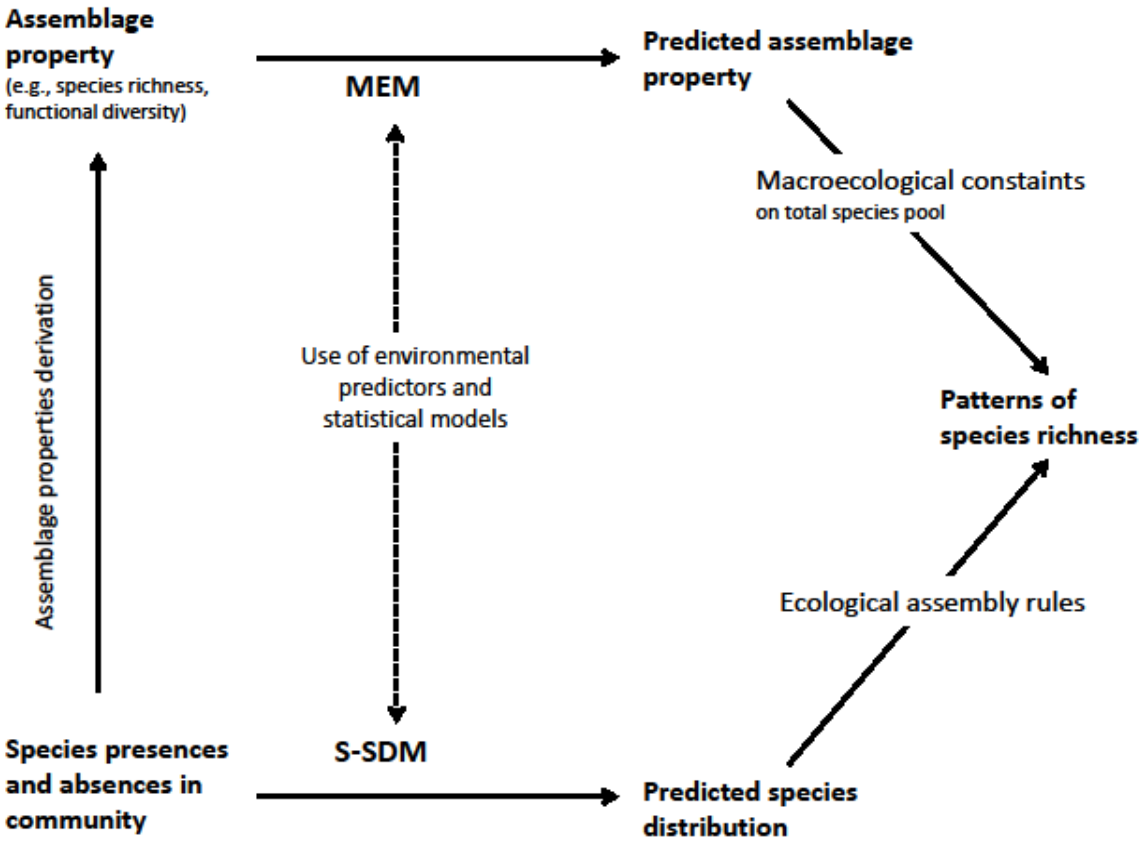


Figure 1. Illustration of the SESAM framework. The MEMs allow predicting properties of plant assemblages such as species richness. These predictions (macro-ecological constraints) are used to limit the number of species to occur in the assemblages predicted on the other hand by S-SDM. Redrawn from Guisan and Rahbek (2012).

STUDY AREA AND DATA COLLECTION

The main chapters of this thesis use a dataset collected in the Swiss Western Alps. Here I present briefly the main characteristics of this study area and the data (environmental data, species occurrences and plant functional traits) that were collected before and during my thesis work.

Localisation, climate and characteristics

The study area is located in the western Alps of Canton de Vaud in Switzerland (Figure 2). It covers ca. 700 km^2 with elevation ranging from 375 m to 3210 m. The climate of this area is temperate: annual temperatures and precipitations fluctuate from 8 °C and 1200 mm at 600 m to -5 °C and 2600 mm at 3000 m. The soil parent material is mainly calcareous. Our studies are focused on open vegetation only (i.e.: meadows, pastures, rocks, screes). Vegetation has long been, and still is, under the influence of human land use. Pastures and meadows interspersed within forest patches are common in this region from the valley bottoms up to the subalpine and lower alpine areas.

Land use is particularly intense at lower altitudes, where the open areas are intensively used as fertilised arable land and grasslands, and are dominated by *Arrhenatherum elatius* and *Poa sp*. At intermediate elevation fertilised pastures and meadows, with *Cynosurus cristatus, Trisetum flavescens, Heracleum sphondylium* are typical. Around the treeline, fertilisation is limited and management is restricted to summer grazing in low productive pastures dominated by *Poa alpina*, *Sesleria caerulea* or *Nardus stricta*.

Alpine belt is characterised with heath, meadow and grassland vegetation (e.g. *Sesleria caerula*, *Nardus stricta*), and finally, the nival belt contain sparse vegetation dominated by taxa typical of high altitudes (e.g. *Saxifraga oppositifolia*).

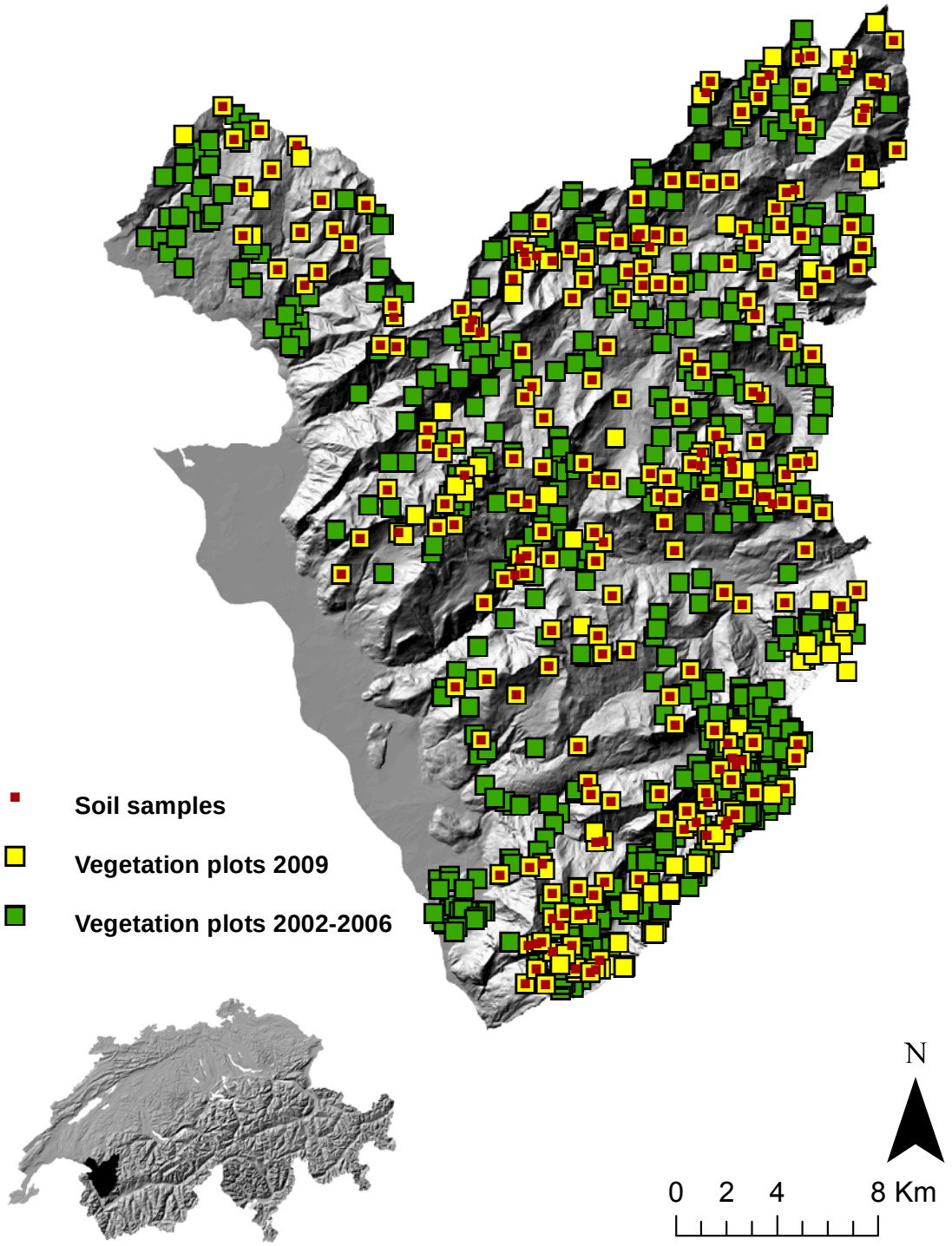


Figure 2. Study area, located in the Western Alps of Switzerland. The vegetation plots sampled between 2002 and 2006 (Modiplant project) are displayed in green, the plots sampled in 2009 (Bioassemble project) are displayed in yellow with a red mark if a soil sample was taken at the same place for chemical analysis.

Vegetation sampling

Our studies focus on open vegetation only (i.e.: meadows, pastures, rocks, screes, see Figure 4). Before the start of the Bioassemble project, 613 4 m^2 exhaustive vegetation plots had already been inventoried in the study area between 2002 and 2006 (Modiplant project, Randin et al., 2009, Figure 2 , green squares). During the summer 2009, we completed this dataset by sampling 299 new vegetation plots following the same random stratified sampling (Figure 2, yellow squares). We used elevation, slope and aspect to stratify the area before sampling and selected an equal number of plots in each strata. This strategy is particularly suited to gather data for species distribution models (Hirzel & Guisan, 2002). On the field, each plot was separated at least by 200m from the others, to avoid spatial autocorrelation.

Plant traits data

During the summer 2009 and 2010, four functional traits were collected for the 240 most abundant species in the study area. We measured data on the vegetative height, (the distance between the top of photosynthetic tissue and the ground), specific leaf area calculated as the ratio of leaf surface to its dry mass, leaf dry matter content calculated as the ratio of leaf dry mass to its saturated fresh mass and the leaves percentage of carbon and nitrogen. Seed mass data were collected from different data bases as well as measured directly (Pellissier et al., 2010b).

Environmental data collected on the field

During the summer 2009, we collected soil samples in 252 vegetation plots to measure edaphic parameters (Figure 2, red squares). For each plot, five samples were taken in the first 10 cm of the organo-mineral horizon after removing the organic horizon and mixed to equalize eventual intra-plot variation. These samples were air-dried, sieved and analyzed later in the lab for phosphorus (standard method), nitrogen and carbon content (elementary analysis with a CHN analyzer), as well as water pH and mineral texture in five classes (clay, coarse and fine sand, coarse and fine silt content).

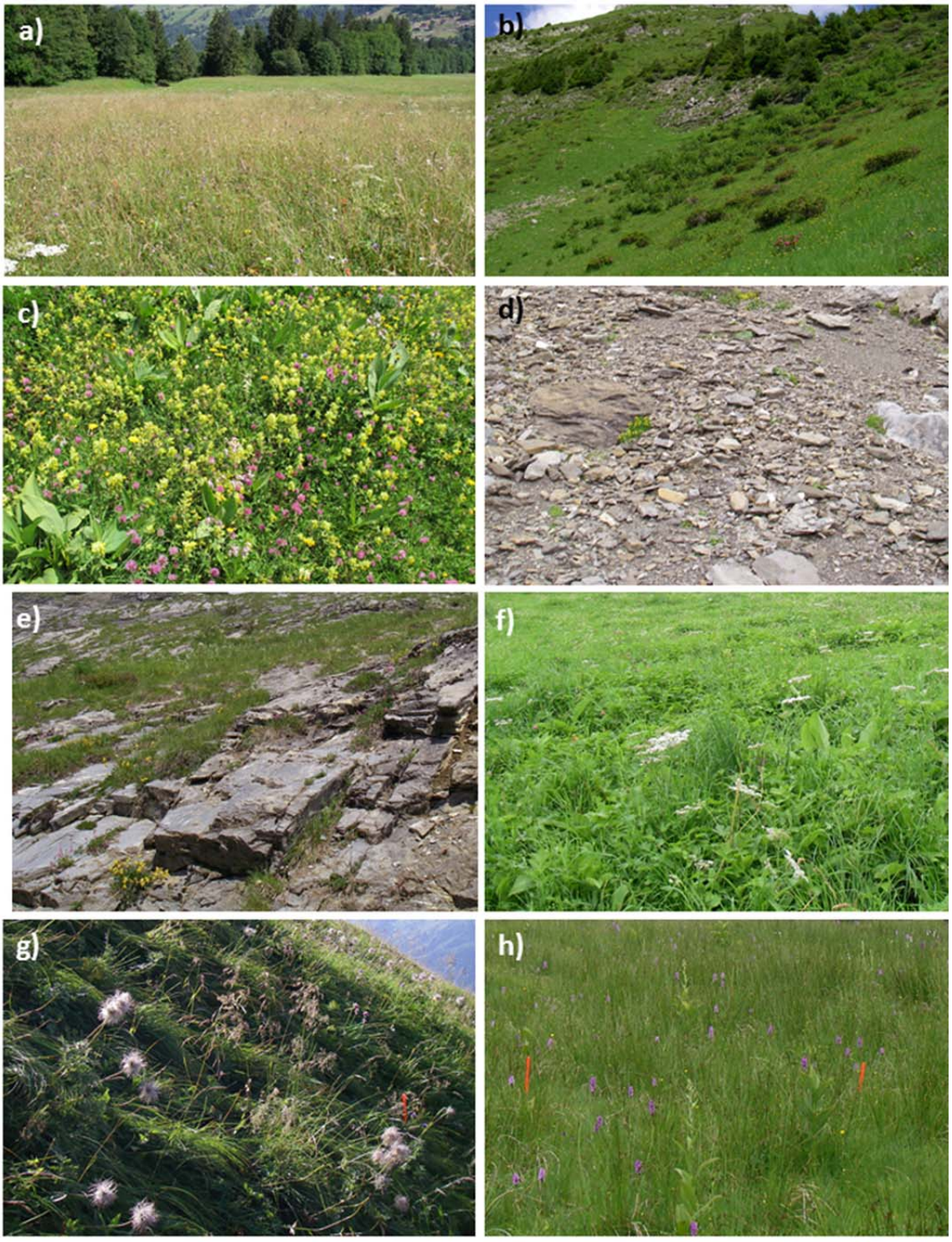


Figure 3. Example of herbaceous vegetation sampled during this thesis, a) mown low elevation grassland, b) meadow on acidic soil, c) calcareous dry meadow, d) calcareous scree vegetation, e) calcareous rocky meadow, f) fertilized low elevation pasture, g) high elevation calcareous meadow on northern slope, h) moist meadow.

OBJECTIVES OF THE THESIS

The objectives of my thesis were to ameliorate SDMs, develop and test MEM to finally combine them to conduct a first test of the SESAM framework. This required several steps: improving SDMs with missing environmental variables (chapter 1), investigating the quality of environmental variables derived from high resolution digital elevation model to be used as predictors in SDMs (chapter 2), comparing MEM and individual species stacking approaches to predicting species richness (chapter 3), developing MEM to predict spatial patterns of functional traits (chapter 4), and finally testing various environmental constraints on S-SDMs to test a first implementation of SESAM (chapter 5).

In **chapter 1**, I tested the importance of seven edaphic variables as predictors in topo-climatic SDMs to assess which one could improve single species prediction as well as assemblage composition, and which had the highest importance as predictor in the models. I also assessed if these predictors had different importance for species with different ecological characteristics.

In **chapter 2**, I tested the performances of environmental variables derived from very high resolution digital elevation model as predictor in SDMs. I compared different grain size of predictors and their impact on the predictive performance of SDMs and S-SDMs.

In **chapter 3**, I tested the respective performances of MEM and S-SDM to predict plant species richness patterns in space and along an elevation gradient. I implemented both approaches on the same dataset and compared their results.

In **chapter 4**, I tested if functional attributes of plant communities (functional traits values and diversity) could be predicted in space and time with regression techniques, what environmental variables were important to explain and predict these patterns. I also projected the resulting models under scenarios of climate change.

Finally in **chapter 5** I used the outputs of the models for species richness and functional traits developed in chapter 3 and 4, to limit the number of species predicted by S-SDM. I tested if this approach allowed predicting a realist number of species and a composition as exact as possible.

The annexes contain papers originating from different project on which I had the opportunity to collaborate during my PhD and that can be related with the chapters presented above.

Annex 1: **Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants**, Pellissier et al. (2010) *Ecography.* In this paper we integrate the distribution of a dominant tundra plant species to SDMs for other plant, and study its impact on the resulting predictions.

Annex 2: **The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients**, Pottier et al. (2013) *Global Ecology and Biogeography*. This paper assesses the ability of topo-climatic S-SDMs to predict plant assemblages along an elevation gradient.

Annex 3: **Climate-based empirical models show biased predictions of butterfly communities along environmental gradients**, Pellissier et al. (2012) *Ecography*. In this paper we test the ability of SSDMs to predict butterfly communities along elevation and plant species richness gradients.

Annex 4: **Ecological assembly rules in plant communities – approaches, patterns and prospects**, Götzenberger et al. (2012) *Biolgical Reviews*. We reviewed papers searching for assembly rules in plant communities by comparing observed patterns with null models simulating random patterns of species assembly.

Annex 5: **A better understanding of ecological conditions for *Leontopodium alpinum* Cassini in the Swiss Alps**, Ischer et al. (in review) *Folia Geobotanica*. In this paper we use species distribution models and phytosociological analysis to define the ecological requirement of the edelweiss.

Finally, I also contributed to the following papers:

Pellissier, L., Pottier, J., Vittoz, P., Dubuis, A., & Guisan, A. (2010). Spatial pattern of floral morphology : possible insight into the effects of pollinators on plant distributions. *Oikos*, *119* (11), pp. 1805-1813.

Pradervand J.N., Pellissier L., Rossier L., Dubuis A., Guisan A., Cherix D. (2011). Diversity of bumblebees (Bombus Latreille, Apidae) in the Alps of the canton Vaud (Switzerland). *Mitteilungen der Schweizerischen Entomologischen Gesellschaft* 84 (1-2) pp. 45-66.

Pellissier L., Fiedler K., Ndribe C., Dubuis A., Pradervand J.N., Guisan A., Rasmann S. (2012). Shifts in species richness, herbivore specialisation and plant resistance along elevation gradients. *Ecology and Evolution* 2 (8) pp. 1818-1825.

Pellissier L., Litsios G., Fiedler K., Pottier J., Dubuis A., Pradervand J.N., Salamin N., Guisan A. (2012). Loss of interactions with ants under cold climate in a regional myrmecophilous butterfly fauna. *Journal of Biogeography* 39(10) pp. 1782-1790.

Pellissier L., Rasmann S., Litsios G., Fiedler K., Dubuis A., Pottier J., Guisan A. (2012). High host-plant nitrogen content: a prerequisite for the evolution of ant-caterpillar mutualism? *Journal of Evolutionary Biology* 25 pp. 1658-1666.

Maiorano L., Cheddadi R., Zimmermann N.E., Pellissier L., Petitpierre B., Pottier J., Laborde H., Hurdu B.I., Pearman P.B., Psomas A., Singarayer J.S., Broennimann O., Vittoz P., Dubuis A., Edwards M.E., Binney H.A., Guisan A. (In Press). Building the niche through time: using 13000 years of data to predict the effects of climate change on three tree species in Europe. *Global Ecology and Biogeography*.

Pellissier L., Alvarez N., Espíndola A., Pottier J., Dubuis A., Pradervand J.N., Guisan A. (In Press). Phylogenetic alpha and beta diversities of butterfly communities correlate with climate in the western Swiss Alps. *Ecography*.

Pellissier L., Anzini M., Maiorano L., Dubuis A., Pottier J., Vittoz P., Guisan A. (In Press). Spatial predictions of land use transitions and associated threats to biodiversity: the case of forest regrowth in mountain grasslands. *Applied Vegetation Science.*

Pellissier L., Ndiribe C., Dubuis A., Pradervand JN., Salamin N., Guisan A., Rasmann S. (In Press). Turnover of plant lineages shapes herbivore phylogenetic beta diversity along ecological gradients. *Ecology Letters.*

Figure 4. Typical landscape in the Western Swiss Alps (Lac Lioson). Open grasslands alternate with forests.

REFERENCES

Adler P.B., Hillerislambers J., & Levine J.M. (2007) A niche for neutrality. *Ecology Letters*, **10**, 95–104.

Akçakaya H.R., Radeloff V.C., Mladenoff D.J., & He H.S. (2004) Integrating Landscape and Metapopulation Modeling Approaches : Viability of the Sharp-tailed Grouse in a Dynamic Landscape. *Conservation Biology*, **18**, 526–537.

Algar A.C., Kharouba H.M., Young E.R., & Kerr J.T. (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography*, **32**, 22–33.

Allouche O., Tsoar A., & 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.

Aranda S.C. & Lobo J.M. (2011) How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. *Ecography*, **34**, 31–38.

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

Araújo M.B., New M., & Araujo M.B. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.

Austin M. (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–19.

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

Barbet-Massin M., Jiguet F., Albert C.H., & Thuiller W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, **3**, 327–338.

Baselga A. & Araújo M.B. (2009) Individualistic vs community modelling of species distributions under climate change. *Ecography*, **32**, 55–65.

Baselga A. & Araújo M.B. (2010) Do community-level models describe community variation effectively? *Journal of Biogeography*, **37**, 1842–1850.

De Bello F., Price J.N., Münkemüller T., Liira J., Zobel M., Thuiller W., Gerhold P., Götzenberger L., Lavergne S., Lepš J., Zobel K., & Pärtel M. (2012) Functional species pool framework to test for biotic effects on community assembly. *Ecology*, **93**, 2263–2273.

De Bello F., Thuiller W., Lepš J., Choler P., Clément J.-C., Macek P., Sebastià M.-T., Lavorel S., Leps J., Clement J.C., & Sebastia M.T. (2009) Partitioning of functional diversity reveals the scale and extent of trait convergence and divergence. *Journal of Vegetation Science*, **20**, 475–486.

Bergamini A., Stofer S., Bolliger J., & Scheidegger C. (2007) Evaluating macrolichens and environmental variables as predictors of the diversity of epiphytic microlichens. *The Lichenologist*, **39**, 475–489.

Boitani L., Sinibaldi I., Corsi F., Biase A., Carranza I. d'Inzillo, Ravagli M., Reggiani G., Rondinini C., & Trapanese P. (2007) Distribution of medium- to large-sized African mammals based on habitat suitability models. *Biodiversity and Conservation*, **17**, 605–621.

Bonan G.B., Levis S., Sitch S., Vertenstein M., & Oleson K.W. (2003) A dynamic global vegetation model for use with climate models: concepts and description of simulated vegetation dynamics. *Global Change Biology*, **9**, 1543–1566.

Breiman L. (2001) Random forests. *Machine Learning*, **45**, 5–32.

Brooker R.W., Callaway R.M., Cavieres L. a, Kikvidze Z., Lortie C.J., Michalet R., Pugnaire F.I., Valiente-Banuet A., & Whitham T.G. (2009) Don't diss integration: a comment on Ricklefs's disintegrating communities. *The American naturalist*, **174**, 919–927.

Bugmann H. (2001) A review of forest gap models. *Climatic change*, **51**, 259–305.

Buisson L., Thuiller W., Casajus N., Lek S., & Grenouillet G. (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145–1157.

Busby J.R. (1991) BIOCLIM a bioclimate analysis and prediction system. *Nature conservation: cost effective biological surveys and data analysis* (ed. by C.R. Margules and M.P. Austin), pp. 64–68.

Cardinale B.J., Duffy J.E., Gonzalez A., Hooper D.U., Perrings C., Venail P., Narwani A., Mace G.M., Tilman D., Wardle D. a, Kinzig A.P., Daily G.C., Loreau M., Grace J.B., Larigauderie A., Srivastava D.S., & Naeem S. (2012) Biodiversity loss and its impact on humanity. *Nature*, **486**, 59–67.

Carroll C. (2010) Role of climatic niche models in focal-species-based conservation planning: Assessing potential effects of climate change on Northern Spotted Owl in the Pacific Northwest, USA. *Biological Conservation*, **143**, 1432–1437.

Carvalho S.B., Brito J.C., Pressey R.L., Crespo E., & Possingham H.P. (2010) Simulating the effects of using different types of species distribution data in reserve selection. *Biological Conservation*, **143**, 426–438.

Chave J. (2004) Neutral theory and community ecology. *Ecology Letters*, **7**, 241–253.

Chen I.-C., Hill J.K., Ohlemüller R., Roy D.B., & Thomas C.D. (2011) Rapid range shifts of species associated with high levels of climate warming. *Science (New York, N.Y.)*, **333**, 1024–6.

Cianfrani C., Lay G. Le, Maiorano L., Satizábal H.F., Loy A., & Guisan A. (2011) Adapting global conservation strategies to climate change at the European scale: The otter as a flagship species. *Biological Conservation*, **144**, 2068–2080.

Clements F.E. (1916) *Plant succession: an analysis of the development of vegetation.* Washington.

Clements F.E. (1936) Nature and structure of the climax. *Journal of Ecology*, **24**, 252– 284.

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

Cornwell W.K. & Ackerly D.D. (2009) Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, **79**, 109–126.

Cornwell W.K., Schwilk D.W., Ackerly D.D., & Schwilk L.D.W. (2006) A trait-based test for habitat filtering: convex hull volume. *Ecology*, **87**, 1465–1471.

Coudun C. & Gegout J.C. (2007) Quantitative prediction of the distribution and abundance of Vaccinium myrtillus with climatic and edaphic factors. *Journal of Vegetation Science*, **70**, 580.

Van Couwenberghe R., Collet C., Pierrat J.-C., Verheyen K., & Gégout J.-C. (2013) Can species distribution models be used to describe plant abundance patterns? *Ecography*, in press.

Davis A.P., Gole T.W., Baena S., & Moat J. (2012) The Impact of Climate Change on Indigenous Arabica Coffee (Coffea arabica): Predicting Future Trends and Identifying Priorities. *PLoS ONE*, **7**, e47981.

Diaz S. & Cabido M. (2001) Vive la différence: plant functional diversity matters to ecosystem processes. *Trends in Ecology & Evolution*, **16**, 646–655.

Dirnböck T., Dullinger S., Grabherr G., & Dirnb T. (2003) A regional impact assessment of climate and land-use change on alpine vegetation. *Journal of Biogeography*, **101**, 1–417.

Douma J.C., Witte J.-P.M., Aerts R., Bartholomeus R.P., Ordoñez J.C., Venterink H.O., Wassen M.J., & Van Bodegom P.M. (2012) Towards a functional basis for predicting vegetation patterns; incorporating plant traits in habitat distribution models. *Ecography*, **35**, 294–305.

Dullinger S., Kleinbauer I., Pauli H., Gottfried M., Brooker R., Nagy L., Theurillat J.-P., Holten J.I., Abdaladze O., Benito J.-L., Borel J.-L., Coldea G., Ghosn D., Kanka R., Merzouki a., Klettner C., Moiseev P., Molau U., Reiter K., Rossi G., Stanisci a., Tomaselli M., Unterlugauer P., Vittoz P., & Grabherr G. (2007) Weak and variable relationships between environmental severity and small-scale co-occurrence in alpine plant communities. *Journal of Ecology*, **95**, 1284–1295.

Elith J., Graham C.H., Anderson R.P., Dudı 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., Williams S., Wisz M.S., Zimmermann N.E., Dudik M., & Soberon J. (2006) Novel methods improve prediction of species ' distributions from occurrence data. *Ecography*, **29**, 129–151.

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.

Engler R., Randin C.F., Thuiller W., Dullinger S., Zimmermann N.E., Araujo M.B., Pearman P.B., Le Lay G., Piedallu C., Albert C.H., Choler P., Coldea G., De Lamo X., Dirnbock T., Gegout J.C., Gomez-Garcia D., Grytnes J.-A.A., Heegaard E., Hoistad F., Nogues-Bravo D., Normand S., Puscas M., Sebastia M.T., Stanisci A., Theurillat J.-P.P., Trivedi M.R., Vittoz P., Guisan A., Araújo M.B., Dirnböck T., Gégout J.-C.,

Gómez-García D., Høistad F., Nogués-Bravo D., Puşcaş M., & Sebastià M.-T. (2011) 21st century climate change threatens mountain flora unequally across Europe. *Global Change Biology*, **17**, 2330–2341.

Ferrier S. & Guisan A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.

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.

Fortunel C., Garnier E., Joffre R., Kazakou E., Quested H., Grigulis K., Lavorel S., Ansquer P., Castro H., Cruz P., Dolezal J., Eriksson O., Freitas H., Golodets C., Jouany C., Kigel J., Kleyer M., Lehsten V., Leps J., Meier T., Pakeman R., Papadimitriou M., Papanastasis V.P., Quetier F., Robson M., Sternberg M., Theau J.-P.P., Thebault A., Zarovali M., Quétier F., & Thébault A. (2009) Leaf traits capture the effects of land use changes and climate on litter decomposability of grasslands across Europe. *Ecology*, **90**, 598–611.

Franklin J. (2009) *Mapping species distributions: spatial inference and prediction.* Cambridge, UK.

Friedman J.H. (2001) Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, **29**, 1189–1232.

Garnier E., Cortez J., Billès G., Navas M.-L.L., Roumet C., Debussche M., Laurent G., Blanchard A., Aubry D., Bellmann A., Neill C., Toussaint J.-P.P., & Billes G. (2004) Plant Functional Markers Capture Ecosystem Properties During Secondary Succession. *Ecology*, **85**, 2630–2637.

Gause G.F. (1936) *The struggle for existence.* Williams and Wilkins, Baltimore.

Gleason H.A. (1926) The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club*, **53**, 1–20.

Gotelli N.J. (2000) NULL MODEL ANALYSIS OF SPECIES CO-OCCURRENCE PATTERNS. *Ecology*, **81**, 2606–2621.

Gould W. (2000) Remote Sensing of Vegetation, Plant Species Richness, and Regional Biodiversity Hotspots. *Ecological Applications*, **10**, 1861–1870.

Grenouillet G., Buisson L., Casajus N., & Lek S. (2011) Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography*, **34**, 9–17.

Grime J.P. (2006) Trait convergence and trait divergence in herbaceous plant communities : Mechanisms and consequences. *Journal of Vegetation Science*, **17**, 255–260.

Grinnell J. (1917) The niche-relationships of the California Thrasher. *Auk*, **34**, 427–433.

Guisan A., Broennimann O., Engler R., Vust M., Yoccoz N.G., Lehmann A., & Zimmermann N.E. (2006) Using Niche-Based Models to Improve the Sampling of Rare Species. *Conservation Biology*, **20**, 501–511.

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

Guisan A., Theurillat J.P., & Kienast F. (1998) Predicting the potential distribution of plant species in an Alpine environment. *Journal of Vegetation Science*, **9**, 65–74.

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

Guisan A., Weiss S.B., & Weiss A.D. (1999) GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, **143**, 107–122.

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

Götzenberger L., De Bello F., Anne Bråthen K., Davison J., Dubuis A., Guisan A., Lepš J., Lindborg R., Moora M., Pärtel M., Pellissier L., Pottier J., Vittoz P., Zobel K., & Zobel M. (2012) Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biological Reviews*, **87**, 111–127.

Hallstan S., Johnson R.K., Willén E., & Grandin U. (2012) Comparison of classification-then-modelling and species-by-species modelling for predicting lake phytoplankton assemblages. *Ecological Modelling*, **231**, 11–19.

Hanley J.A. & Mcneil B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) Curve. *Radiology*, **143**, 29–36.

Hastie T. & Tibshirani R. (1990) *Generalized Additive Models.* Chapman and Hall, London.

Hector A., Schmid B., Beierkuhnlein C., Caldeira M.C., Diemer M., Dimitrakopoulos P.G., Finn J.A., Freitas H., Giller P.S., Good J., Harris R., Hogberg P., Huss-Danell K., Joshi J., Jumpponen A., Korner C., Leadley P.W., Loreau M., Minns A., Mulder C.P.H., O'Donovan G., Otway S.J., Pereira J.S., Prinz A., Read D.J., Scherer-Lorenzen M., Schulze E.D., Siamantziouras A.S.D., Spehn E.M., Terry A.C., Troumbis A.Y., Woodward F.I., Yachi S., & Lawton J.H. (1999) Plant diversity and productivity experiments in European grasslands. *Science*, **286**, 1123–1127.

Hirzel A. & Guisan A. (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331–341.

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

Hooper D.U., Chapin F.S., Ewel J.J.J., Hector A., Inchausti P., Lavorel S., Lawton J.H., Lodge D.M., Loreau M., Naeem S., Schmid B., Setala H., Symstad A.J.J., Vandermeer J., & Wardle D.A. (2005) Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs*, **75**, 3–35.

Hortal J., De Marco P., Santos A.M.C., & Diniz-Filho J.A.F. (2012) Integrating biogeographical processes and local community assembly. *Journal of Biogeography*, **39**, 627–628.

Hubbell S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography.* Princeton, NJ.

Hubbell S.P. (2005) Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology*, **84**, 3174–172.

Hutchinson G.E. (1957) Population studies - animal ecology and demography - concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

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

Jones M.C., Dye S.R., Pinnegar J.K., Warren R., & Cheung W.W.L. (2012) Modelling commercial fish distributions: Prediction and assessment using different approaches. *Ecological Modelling*, **225**, 133–145.

Küster E.C., Bierman S.M., Klotz S., Kühn I., & and Kühn I. (2011) Modelling the impact of climate and land use change on the geographical distribution of leaf anatomy in a temperate flora. *Ecography*, **34**, 507–518.

Lavorel S., Díaz S., Cornelissen J.H.C., Garnier E., Harrison S.P., Mcintyre S., Pausas J.G., Pérez-Harguindeguy N., Roumet C., & Urcelay C. (2007) Plant Functional Types : Are We Getting Any Closer to the Holy Grail ? *Terrestrial Ecosystems in a Changing World* (ed. by J.. G. Canadell, D. Pataki, and L. Pitelka), Springer-Verlag, Berlin Heidelberg.

Lavorel S. & Garnier E. (2002) Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology*, **16**, 545–556.

Le Lay G., Engler R., Franc E., & Guisan A. (2010) Prospective sampling based on model ensembles improves the detection of rare species. *Ecography*, **33**, 1015–1027.

Leathwick J.R. (1998) Are New Zealand's Nothofagus species in equilibrium with their environment ? *Journal of Vegetation Science*, **9**, 719–732.

Leathwick J.R. & Austin M.P. (2001) Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology*, **82**, 2560–2573.

Leibold M.A. & McPeek M.A. (2006) Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, **87**, 1399–410.

Levins R. (1966) The strategy of model building in population ecology. *American Scientist*, **54**, 421–431.

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

Lomba A., Pellissier L., Randin C., Vicente J., Moreira F., Honrado J., & Guisan A. (2010) Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, **143**, 2647–2657.

Luoto M., Virkkala R., Heikkinen R.K., & Rainio K. (2004) Predicting Bird Species Richness Using Remote Sensing in Boreal Agricultural-Forest Mosaics. *Ecological Applications*, **2**, 587–1962.

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

MacArthur R. & Levins R. (1967) Limiting similarity, convergence, and divergence of coexisting species. *American Naturalist*, **101**, 377–385.

Machado-Machado E.A. (2012) Empirical mapping of suitability to dengue fever in Mexico using species distribution modeling. *Applied Geography*, **33**, 82–93.

Maestre F.T., Escolar C., Martinez I., & Escurdo A. (2008) Are soil lichen communities structured by biotic interactions ? A null model analysis. *Journal of Vegetation Science*, **19**, 261–266.

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

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

Magnin-Gonze J. (2004) *Histoire de la botanique.* Paris.

Maiorano L., Cheddadi R., Zimmermann N.E., Pellissier L., Petitpierre B., Pottier J., Laborde H., Hurdu B.I., Pearman P.B., Psomas A., Singarayer J.S., Broennimann O., Vittoz P., Dubuis A., Edwards M.E., Binney H. a., & Guisan A. (2013) Building the niche through time: using 13,000 years of data to predict the effects of climate change on three tree species in Europe. *Global Ecology and Biogeography*, in press.

Marmion M., Luoto M., Heikkinen R.K., & Thuiller W. (2009) The performance of state-of-the-art modelling techniques depends on geographical distribution of species. *Ecological Modelling*, **220**, 3512–3520.

Martin C.S., Vaz S., Ellis J.R., Lauria V., Coppin F., & Carpentier A. (2012) Modelled distributions of ten demersal elasmobranchs of the eastern English Channel in relation to the environment. *Journal of Experimental Marine Biology and Ecology*, **418-419**, 91–103.

Mayfield M.M. & Levine J.M. (2010) Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology letters*, **13**, 1085–93.

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

Meier E.S., Kienast F., Pearman P.B., Svenning J.-C., Thuiller W., Araújo M.B., Guisan A., & Zimmermann N.E. (2010) Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, **33**, 1038–1048.

Mokany K., Ash J., & Roxburgh S. (2008) Functional identity is more important than diversity in influencing ecosystem processes in a temperate native grassland. *Journal of Ecology*, **96**, 884–893.

Mouchet M. a., Villéger S., Mason N.W.H., Mouillot D., & Villeger S. (2010) Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Functional Ecology*, **24**, 867–876.

Newbold T., Gilbert F., Zalat S., El-Gabbas A., & Reader T. (2009) Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography*, **36**, 2085–2095.

Parviainen M., Marmion M., Luoto M., Thuiller W., & Heikkinen R.K. (2009) Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. *Biological Conservation*, **142**, 2501–2509.

Pearman P.B., Randin C.F., Broennimann O., Vittoz P., Van der Knaap W.O., Engler R., Le Lay G., Zimmermann N.E., & Guisan A. (2008) Prediction of plant species distributions across six millennia. *Ecology letters*, **11**, 357–69.

Pellissier L., Bråthen K.A., Pottier J., Randin C.F., Vittoz P., Dubuis A., Yoccoz N.G., Alm T., Zimmermann N.E., Guisan A., & Brathen K.A. (2010a) Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, **33**, 1004–1014.

Pellissier L., Fournier B., Guisan A., & Vittoz P. (2010b) Plant traits co-vary with altitude in grasslands and forests in the European Alps. *Plant Ecology*, **211**, 351–365.

Pellissier L., Pottier J., Vittoz P., Dubuis A., & Guisan A. (2010c) Spatial pattern of floral morphology : possible insight into the effects of pollinators on plant distributions. *Oikos*, **119**, 1805–1813.

Peppler-Lisbach C., Schröder B., & Schroder B. (2004) Predicting the species composition of Nardus stricta communities by logistic regression modelling. *Journal of Vegetation Science*, **15**, 623–634.

Pereira H.M., Leadley P.W., Proença V., Alkemade R., Scharlemann J.P.W., Fernandez-Manjarrés J.F., Araújo M.B., Balvanera P., Biggs R., Cheung W.W.L., Chini L., Cooper H.D., Gilman E.L., Guénette S., Hurtt G.C., Huntington H.P., Mace G.M., Oberdorff T., Revenga C., Rodrigues P., Scholes R.J., Sumaila U.R., & Walpole M. (2010) Scenarios for global biodiversity in the 21st century. *Science (New York, N.Y.)*, **330**, 1496–501.

Petitpierre B., Kueffer C., Broennimann O., Randin C., Daehler C., & Guisan A. (2012) Climatic niche shifts are rare among terrestrial plant invaders. *Science (New York, N.Y.)*, **335**, 1344–8.

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

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

Pinto P.E., Gégout J., Hervé J., & Dhôte J. (2007) Changes in environmental controls on the growth of Abies alba Mill. in the Vosges Mountains , north-eastern France , during the 20th century. *Global Ecology and Biogeography*, **16**, 472–484.

Pio D. V., Broennimann O., Barraclough T.G., Reeves G., Rebelo A.G., Thuiller W., Guisan A., & Salamin N. (2011) Spatial Predictions of Phylogenetic Diversity in Conservation Decision Making. *Conservation Biology*, **25**, 1229–1239.

Pompe S., Hanspach J., Badeck F., Klotz S., Thuiller W., & Kuhn I. (2008) Climate and land use change impacts on plant distributions in Germany. *Biology Letters*, **4**, 564–567.

Pontes L.D.S., Soussana J.-F.F., Louault F., Andueza D., Carrere P., & Carrère P. (2007) Leaf traits affect the above-ground productivity and quality of pasture grasses. *Functional Ecology*, **21**, 844–853.

Pottier J., Dubuis A., Pellissier L., Maiorano L., Rossier L., Randin C.F., Vittoz P., & Guisan A. (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, **22**, 52–63.

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

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

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

Randin C.F., Jaccard H., Vittoz P., Yoccoz N.G., & Guisan A. (2009b) Land use improves spatial predictions of mountain plant abundance but not presence-absence. *Journal of Vegetation Science*, **44**, 996–1008.

Randin C.F., Vuissoz G., Liston G.E., Vittoz P., & Guisan A. (2009c) Introduction of Snow and Geomorphic Disturbance Variables into Predictive Models of Alpine Plant Distribution in the Western Swiss Alps. *Arctic, Antarctic, and Alpine Research*, **41**, 347–361.

Ricklefs R.E. (2008) Disintegration of the ecological community. *The American naturalist*, **172**, 741–50.

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

Ripley B.D. (1996) *Pattern recognition and neural networks.*

Rondinini C., Di Marco M., Chiozza F., Santulli G., Baisero D., Visconti P., Hoffmann M., Schipper J., Stuart S.N., Tognelli M.F., Amori G., Falcucci A., Maiorano L., & Boitani L. (2011) Global habitat suitability models of terrestrial mammals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **366**, 2633–41.

Le Roux P.C., Lenoir J., Pellissier L., Wisz M.S., & Luoto M. (2013) Horizontal, but not vertical, biotic interactions affect fine-scale plant distribution patterns in a low energy system. *Ecology*, in press.

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., & Wall D.H. (2000) Global biodiversity scenarios for the year 2100. *Science (New York, N.Y.)*, **287**, 1770–4.

Sanchez A.C., Osborne P.E., & Haq N. (2010) Identifying the global potential for baobab tree cultivation using ecological niche modelling. *Agroforestry Systems*, **80**, 191–201.

Schleuter D., Daufresne M., Massol F., & Argillier C. (2010) A user's guide to functional diversity indices. *Ecological Monographs*, **80**, 469–484.

Schorr G., Holstein N., Pearman P.B., Guisan A., & Kadereit J.W. (2012) Integrating species distribution models (SDMs) and phylogeography for two species of Alpine Primula. *Ecology and evolution*, **2**, 1260–77.

Shipley B., Vile D., & Garnier E. (2006) From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science*, **314**, 812–14.

Sitch S., Smith B., Prentice I.C., Arneth A., Bondeau A., Cramer W., Kaplan J.O., Levis S., Lucht W., Sykes M.T., Thonicke K., & Venevsky S. (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, **9**, 161–185.

Soberón J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology letters*, **10**, 1115–1123.

Sonnier G., Shipley B., & Navas M.L. (2010) Quantifying relationships between traits and explicitly measured gradients of stress and disturbance in early successional plant communities. *Journal of Vegetation Science*, **21**, 318–331.

Spehn E.M., Hector A., Joshi J., Scherer-Lorenzen M., Schmid B., Bazeley-White E., Beierkuhnlein C., Caldeira M.C., Diemer M., Dimitrakopoulos P.G., Finn J.A., Freitas H., Giller P.S., Good J., Harris R., Hogberg P., Huss-Danell K., Jumpponen A., Koricheva J., Leadley P.W., Loreau M., Minns A., Mulder C.P.H., O'Donovan G., Otway S.J., Palmborg C., Pereira J.S., Pfisterer A.B., Prinz A., Read D.J., Schulze E.D., Siamantziouras A.S.D., Terry A.C., Troumbis A.Y., Woodward F.I., Yachi S., & Lawton J.H. (2005) Ecosystem effects of biodiversity manipulations in European grasslands. *Ecological Monographs*, **75**, 37–63.

Steinmann K., Linder H.P., & Zimmermann N.E. (2009) Modelling plant species richness using functional groups. *Ecological Modelling*, **220**, 962–967.

Stokes C.J. & Archer S.R. (2010) Niche differentiation and neutral theory: an integrated perspective on shrub assemblages in a parkland savanna. *Ecology*, **91**, 1152–1162.

Sundblad G., Härmä M., Lappalainen A., Urho L., & Bergström U. (2009) Transferability of predictive fish distribution models in two coastal systems. *Estuarine, Coastal and Shelf Science*, **83**, 90–96.

Tang J. & Zhou S. (2012) Hybrid niche-neutral models outperform an otherwise equivalent neutral model for fitting coral reef data. *Journal of theoretical biology*, **317C**, 212–218.

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

Thuiller W., Richardson D.M., Pysek P., Midgley G.F., Hughes G.O., & Rouget M. (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, **11**, 2234–2250.

Titeux N., Maes D., Marmion M., Luoto M., & Heikkinen R.K. (2009) Inclusion of soil data improves the performance of bioclimatic envelope models for insect species distributions in temperate Europe. *Journal of Biogeography*, **36**, 1459–1473.

Tognelli M.F., Abba A.M., Bender J.B., & Seitz V.P. (2010) Assessing conservation priorities of xenarthrans in Argentina. *Biodiversity and Conservation*, **20**, 141–151.

Vega R., Fløjgaard C., Lira-Noriega A., Nakazawa Y., Svenning J.-C., & Searle J.B. (2010) Northern glacial refugia for the pygmy shrew Sorex minutus in Europe revealed by phylogeographic analyses and species distribution modelling. *Ecography*, **33**, 260–271.

Venn S.E., Green K., Pickering C.M., & Morgan J.W. (2011) Using plant functional traits to explain community composition across a strong environmental filter in Australian alpine snowpatches. *Plant Ecology*, **212**, 1491–1499.

Vicente J., Randin C.F., Goncalves J., Metzger M.J., Lomba A., Honrado J., & Guisan A. (2011) Where will conflicts between alien and rare species occur after climate and land-use change? A test with a novel combined modelling approach. *Biological Invasions*, **13**, 1209–1227.

Violle C., Navas M.L., Vile D., Kazakou E., Fortunel C., Hummel I., & Garnier E. (2007) Let the concept of trait be functional ! *Oikos*, **116**, 882–892.

Virtanen R., Oksanen J., Oksanen L., Razzhivin V.Y., & Yu V. (2006) Broad-scale vegetation-environment relationships in Eurasian high-latitude areas. *Journal of Vegetation Science*, **17**, 519–528.

Volterra V. (1926) Fluctuations in the abundance of a species considered mathematically. *Nature*, **118**, 196–218.

Weiher E. & Keddy P. (1999) *Ecological assembly rules: perspectives, advances, retreats.* Cambridge, UK.

Wennekes P.L., Rosindell J., & Etienne R.S. (2012) The Neutral-Niche Debate: A Philosophical Perspective. *Acta biotheoretica*, **60**, 257–271.

Wiens J. a, Stralberg D., Jongsomjit D., Howell C. a, & Snyder M. a (2009) Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19729–19736.

Wilson J.B. (2007) Trait-divergence assembly rules have been demonstrated: Limiting similarity lives! A reply to Grime. *Journal of Vegetation Science*, **18**, 451–452.

Wilson J.B., Gitay H., & Agnew A.D.Q. (1987) Does niche limitation exist ? *Functional Ecology*, **1**, 391–397.

Wilson J.B. & Stubbs W.J. (2012) Evidence for assembly rules: limiting similarity within a saltmarsh. *Journal of Ecology*, **100**, 210–221.

Wisz M.S., Walther B. a., & Rahbek C. (2007) Using potential distributions to explore determinants of Western Palaearctic migratory songbird species richness in sub-Saharan Africa. *Journal of Biogeography*, **34**, 828–841.

Wohlgemuth T. (1998) Modelling floristic species richness on a regional scale : a case study in Switzerland. *Biodiversity and Conservation*, **7**, 159–177.

Yee T.W. & Mackenzie M. (2002) Vector generalized additive models in plant ecology. *Ecological Modelling*, **157**, 141–156.

CHAPTER 1

Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables

Anne Dubuis [1]
Sara Giovanettina [1]
Loïc Pellissier [1]
Julien Pottier [2]
Pascal Vittoz [1]
Antoine Guisan [1,3]

1 Department of Ecology and Evolution, University of Lausanne, Biophore building, 1015 Lausanne, Switzerland

2 INRA, Grassland Ecosystem Research Unit (UREP), 5 Chemin de Beaulieu, 63100 Clermont-Ferrand, France

3 Faculty of Geoscience and the Environment, University of Lausanne, 1015 Lausanne, Switzerland

This paper is in press in Journal of Vegetation Science
DOI: 10.1111/jvs.12002

Contribution to the project: I participated to the field work for plant, functional traits and soil data collection. I did soil chemical analysis in collaboration with SG, I performed the species distribution modelling and analysed the results. I also wrote the initial draft of the paper and lead the reviewing process.

ABSTRACT

Questions: Soil properties have been widely shown to influence plant growth and distribution. However, the degree to which edaphic variables can improve models based on topo-climatic variables is still unclear. In this study, we tested the roles of seven edaphic variables, namely (1) pH, the content of (2) nitrogen and of (3) phosphorus, (4) silt, (5) sand and (6) clay and the (7) carbon-to-nitrogen ratio, as predictors in species distribution models in an edaphically heterogeneous landscape. We also tested how the respective influence of these variables in the models is linked to different ecological and functional species characteristics.

Location: The Western Alps of Switzerland

Methods: With four different modelling techniques, we built models for 115 plant species using topo-climatic variables only and then topo-climatic variables plus each of the seven edaphic variables one at a time. We evaluated the contribution of each edaphic variable by assessing the change in the predictive power of the model. In a second step we evaluated the importance of the two edaphic variables that yielded the greatest increase in predictive power in one final set of models for each species. Third, we explored the change in predictive power and the importance of variables across plant functional groups. Finally, we assessed the influence of the edaphic predictors on the prediction of community composition by stacking the models for all species and comparing the predicted communities with the observed community.

Results: Among the set of edaphic variables studied, pH and nitrogen content showed the highest contributions to the improvement of the predictive power of the models, as well as the predictions of community composition. When considering all topo-climatic and edaphic variables together, pH was the second most important variable after degree-days. The changes in the model results caused by edaphic predictors were dependent on species characteristics. The predictions for the species that have a low specific leaf area, and acidophilic preferences, tolerating low soil pH and high humus content, showed the greatest improvement by the addition of pH and nitrogen in the model.

Conclusions: pH was an important predictor variable for explaining the species distribution and community composition of the mountain plants considered in our study. pH allowed more precise predictions for acidophilic species. This variable should not be neglected in the construction of species distribution models in areas with contrasting edaphic conditions.

Keywords: acidophilic species; mountain flora; soil nitrogen content; soil pH; species distribution models (SDMs); Western Swiss Alps.

Nomenclature: Aeschimann & Heitz (1996)

Abbreviations: A = soil acidity; AUC = area under the curve; CL = clay content; C:N = carbon-to-nitrogen ratio; H = soil humus content; LDMC = leaf dry matter content; N = nitrogen content; Nu = soil nutrient content; P = phosphorus content; Sa = sand content; SDM = species distribution model; Si = silt content; SLA = specific leaf area; TC = topo-climatic; TSS = true skill statistics; VH = vegetative height.

Running head: Edaphic variable importance for plant species distribution models.

INTRODUCTION

The ecological preferences of plant species for particular soil properties are known to influence plant distributions. For instance, Alvarez et al. (2009) demonstrated that ecological preferences for different bedrock types and thus, soil acidity, determined the primary patterns of refugia and plant migrations during the last glacial era and the subsequent climate variations of the Holocene. Similarly, the manuring of agricultural grasslands within the past decades has induced profound changes in their floristic compositions (e.g., Zechmeister et al., 2003; Peter et al., 2008).

The success of a plant is largely conditioned by the soil's chemical properties. These properties determine ion concentration and soil structure and are important for the soil's capacity to hold water. Ions contained in the soil (e.g., nitrogen, calcium or phosphorus) are used as nutrients by the plants (Aerts & Chapin, 2000). Hence, variations in ionic concentrations in the soil directly impact plant growth and, consequently, the formation of plant communities. To a large extent, the availability of ions depends on soil acidity, making soil pH another important factor influencing plant distributions (Gobat et al., 2004). Low soil acidity (high pH) can prevent the release of important ions, causing nutrient deficiencies in the plant. High soil acidity can also cause deficiencies because elements such as phosphorus or nitrogen can form complexes with other ions and become unavailable to plants (Gobat et al., 2004). Conversely, high soil acidity (low pH) enhances the solubilisation of several elements, especially metals such as aluminium or iron that can be toxic to maladapted plants. Finally, soil texture has a physical impact on plants by supporting or limiting root growth and influencing the amount of water and oxygen available to the plants (Bouma & Bryla, 2000; Martre et al., 2002). The proportion of clay also directly influences the clay-organic complex and the associated soil fertility (Gobat et al., 2004).

Plants vary in their functional traits and thus in their preference towards edaphic conditions. Plants displaying tougher, long-living leaves, which better conserve nutrients in the tissues, are expected to grow better in acidic soils, which are frequently nutrient poor. In contrast, if nutrients are not limiting, plant species that are able to rapidly assimilate nutrients and to convert them into biomass have a competitive advantage over less efficient species (Grime, 1977; Pellissier et al., 2010b). Finally, generalist species, which are able to grow under a wide range of soil conditions, are less affected by edaphic factors.

Thus, soil is an important factor explaining plant distributions. However, the degree to which edaphic variables contribute to species distributions, beyond other environmental variables such as climate or topography, is still unclear. One way to address this question is to ask to what degree the predictive power of species distribution models (SDMs; Guisan & Zimmermann, 2000) can be improved by the addition of edaphic variables as predictors. Although numerous studies have used SDMs during the last decade, few focused on the influence that soil may have on plant distributions. Most of these studies have investigated trees and shrubs (Coudun & Gegout, 2005; Coudun et al., 2006; Coudun & Gegout, 2007; Marage & Gegout, 2009) which show small functional differences between species. However, the importance of soil-derived variables in SDMs will most likely vary widely in accordance with species traits.

We evaluated the importance of seven edaphic variables ((1) the soil pH, its content of (2) nitrogen (N) and of (3) phosphorus (P), (4) sand (Sa), (5) silt (Si) and (6) clay (Cl), and (7) the carbon-to-nitrogen ratio (C:N)) to explain the distributions of 115 mountain grassland plant species with different ecological characteristics (i.e., the ecological indicator values for their preferred soil properties (Landolt et al., 2010)) and functional traits as well as the composition and species richness of the plant communities. More specifically, we explored the extent to which the edaphic variables used as predictors improved the predictive power of the SDMs based on topo-climatic variables, and we determined the characteristics of the species with the most-improved model results. In a second step, we built topo-climatic models for each species containing the most influential edaphic variables to test the relative importance of the variables and related the importance values of the variables to the different ecological and functional characteristics of the species. Finally, we predicted plant communities from the SDM results and assessed the importance of edaphic variables in this context. For these analyses, we used a comprehensive dataset of vegetation plots together with a set of environmental predictors at a fine resolution and a dataset of plant functional traits for all modelled species.

METHODS

Study area

Our study area is located in the Western Swiss Alps (Canton de Vaud, Switzerland, 46°10' to 46°30'N; 6°50' to 7°10'E, Fig. 1). It covers ca. 700 km^2 and has an elevation stretching from 375 m asl to 3210 m asl. The climate is temperate. The annual temperatures and precipitations vary from 8 °C and 1200 mm, respectively, at 600 m to -5 °C and 2600 mm at 3000 m (Bouët, 1985). The bedrock is primarily calcareous. The soils are variable, ranging from deep brown soils to shallow lithological soils, the latter exerting a greater influence on vegetation. Basic cations and carbonates promote an alkaline pedogenesis that can be followed by an acidic pedogenesis after carbonate lixiviation (Gobat et al., 2004). Spaltenstein (1984) demonstrated a locally important post-glacial loess accumulation in this region. The accumulation of loess accelerated the acidic transition of the pedogenesis. Human influence on the vegetation is important. Deforested areas are often used as pastures or mowed, and manuring is important from the lowlands to the subalpine belt. The grasslands in the alpine belt are used as pastures without manuring. For more information on this geographical area, see Randin et al. (2006).

Species data

During the summer of 2009, 252 vegetation plots, 4 m^2 in size, were chosen using a random stratified sampling procedure (Hirzel & Guisan, 2002) based on elevation, slope and aspect and extensively inventoried. The sampling was limited to open, non-woody areas. The vegetation plots cover an altitudinal range from 820 to 2680 m asl (Fig. 1). Each sampling point was separated from the others by a minimum distance of 200 m, as it has been shown that from this distance on, no autocorrelation was observed between plots in this study area (Pottier et al., 2012). We only included the species that occurred in more than 20 plots throughout the dataset for the following modelling analyses (*i.e.,* 115 species, App. S1).

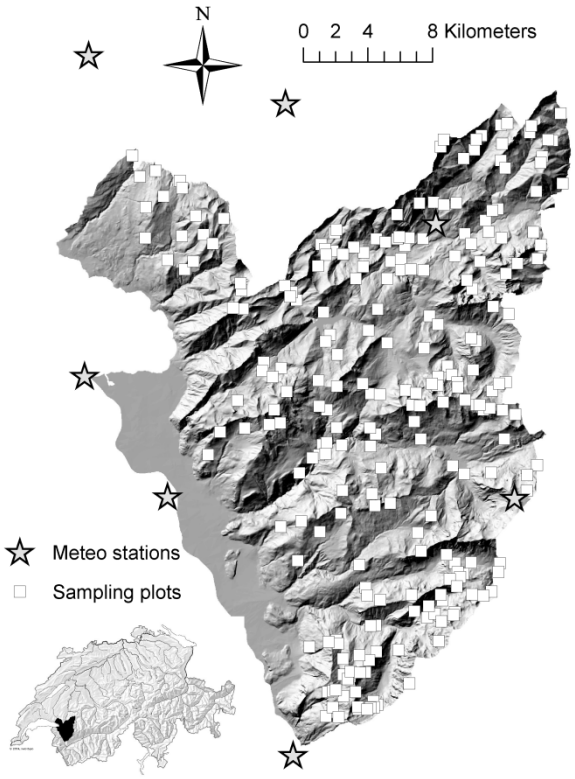


Fig. 1. Map of the study area situated in the western Alps of Canton de Vaud in Switzerland. The 252 plots for which we have edaphic and vegetation data are indicated as white squares, the background shows the relief of the study area, the meteorological stations are displayed as stars.

Climatic and topographic predictors

Three climatic predictors (degree-days, moisture index of the growing season and global solar radiation) and two topographic predictors (slope and topographic position) were used at a resolution of 25 m. The climatic predictors were computed from the monthly means of the average temperature (°C) and sum of precipitation (mm) data recorded for the period 1961-1990 by the Swiss network of meteorological stations (www.meteosuisse.ch). These data are interpolated on Switzerland based on a 25 m resolution digital elevation model (from the Swiss Federal Office of Topography, www.swisstopo.ch) with local thin-plate spline-functions for temperature and regionalized linear regression model for precipitation (Zimmermann & Kienast, 1999). Degree-days are defined as the sum of days of the growing season (June, July and August) multiplied by the temperature above 0°C, see Zimmermann & Kienast (1999) for the formula. The moisture index is defined as the difference between precipitation and potential evapotranspiration and represents the potential amount of water available in the soil at a site. In this study, we use the sum of the mean daily values for the

growing season. The potential global solar radiation was calculated over the year. The methods for computing these predictors are described in more detail in Zimmermann & Kienast (1999). The slope in degrees was derived from the digital elevation model with ArcGis 9.3 spatial analyst tool (ESRI 2008). The topographic position is an integration of topographic features at various scales and is computed with moving windows. Positive values of this variable indicate relative ridges and tops, whereas negative values correspond to valleys and sinks. This variable was calculated using the method from Zimmermann (2007) . These five variables are recognised as theoretically meaningful for explaining plant distributions (Körner, 1999) and have already been used successfully in several modelling studies in the same study area (Engler et al., 2009; Randin et al., 2009c; Pellissier et al., 2010a).

Edaphic predictors

During the vegetation field survey, five soil samples were taken per vegetation plot; one sample was taken from each corner of the plot, and one was taken from the middle. The samples were taken from the top 10 cm of the soil after removing the first organic soil horizon, corresponding primarily to the organo-mineral horizon (Gobat et al., 2004). The samples were eventually mixed to equalize intra-plot variation, air-dried, sieved at 2 mm and ground into powder.

We measured the following soil properties: water pH, total nitrogen and phosphorus content, organic carbon and mineral texture in terms of three classes (clay, silt and sand). The pH was measured with a pH meter after diluting soil in water in a 1:2.5 proportion. The total carbon and nitrogen content were analysed by elementary analysis after combustion with a CHN analyser. We then estimated the amount of organic carbon by removing the mineral carbon (derived by calcimetric analyses with a Bernard calcimeter) according to the methods of Baize (2000), if the soil pH was above 6.5. Phosphorus content was determined by colourimetric analysis after mineralisation at 550°C (Baize, 2000).

Species distribution modelling

A summary of the following analyses can be found in the flowchart presented in Fig. 2. We used the BIOMOD library (Thuiller et al., 2009) in the R software (2.13.0, R Development Core Team) to model the distribution of the 115 plant species using four different modelling techniques (two regression methods and two classification methods): generalised linear models (GLM; McCullagh & Nelder, 1989), generalised additive models (GAM; Hastie & Tibshirani, 1990), generalised boosted models (GBM;

Ridgeway, 1999; Friedman et al., 2000) and random forests (RF; Breiman, 2001). We used four methods as the variability between techniques has been raised in many papers as an important source of uncertainty (Elith et al., 2006; Guisan et al., 2007; Buisson et al., 2010). Ensemble modelling (Araujo & New, 2007), is a way to limit and quantify this uncertainty. The four modelling techniques used in this study were shown to provide satisfactory predictions of species distributions in a large comparative study (Elith et al., 2006). GLM and GAM were calibrated using a binomial distribution and a logistic link function, and polynomials were allowed up to second order (linear and quadratic terms) for each predictor. GBM was calibrated using 2000 trees and RF with default parameters from BIOMOD. We calibrated eight sets of models for each species, one set of models using only the topo-climatic predictors and one set of models for each of the seven edaphic predictors (pH, N, P, Sa, Si, Cl and C:N) in addition to the five topo-climatic (TC) predictors. The distribution of all predictors is presented in App. S5, and the correlation between them in App. S3. The models were validated with a repeated (10 times) random split-sample procedure using 70% of the data points to fit the models and 30% for an independent validation. For each data partition, the predictive power was estimated with the Area Under the Curve (AUC) of a receiver operating characteristic plot (Fielding & Bell, 1997) and the true skill statistic (TSS; Allouche et al., 2006). The AUC values range from 0.5 for models with random predictions to 1 for models perfectly fitting the data. A model is rated as fair if its AUC is higher than 0.7 (Swets, 1988). The TSS values vary between 0 for a random model and 1 for a model showing perfect agreement.

Predictive power analysis

We evaluated the global contributions of each edaphic variable to the models of our species by focusing on the mean change in the predictive power (AUC or TSS) of all four modelling techniques caused by the addition of the respective edaphic variable to the models. We used the mean of the change across the four techniques, instead of relying on one modelling technique alone to draw conclusion on variables contributions in the models. Using the mean changes is a way to consider only changes that are consistent through modelling techniques. To verify whether the different modelling techniques showed a comparable variation due to the addition of edaphic predictors, we measured the correlations of the predictive power change due to the addition of each edaphic predictor for each pair of modelling techniques. We then calculated the percentage of species with improved or worsened models. To discard species for which the change in the AUC was too small, we only considered the predictive power

of the model to be improved or worsened if the AUC difference exceeded 0.02 (ad-hoc threshold). Previous studies had simply quantified any improvement or degradation of the AUC (Randin et al., 2009b; Randin et al., 2009a), so our approach is comparatively more conservative in the sense that our estimates of improvement are, at worst underestimated.

To gain a better understanding regarding the ecological implications of the results, we explored the changes in predictive power for the species after grouping them according to three categorical ecological characteristics. We used three ecological indicator values (Landolt et al., 2010) to classify our species according to their predicted tolerance to three different soil characteristics: the soil pH, nutrient content and humus value. The ecological values for soil pH ranged from 1, for species with tolerance to low pH, to 5, for species growing in base-saturated soil. The nutrient content values ranged from 1, for species that grow on poor soils, to 5, for species favouring nutrient-rich soil. The humus content values ranged from 1, for species growing on humus-poor soil, to 5, for species preferring humus-rich soil. We also considered the predictive power changes as a function of three continuous functional traits using linear regression The three functional traits for which we had collected field measurements on most of the modelled species are leaf dry matter content (LDMC), specific leaf area (SLA) and plant vegetative height (VH). LDMC, SLA and VH were averaged from measurements of 10 individuals of each species, all sampled within the study area (see Pottier et al. (2012) for more information). All ecological characteristics and trait values are presented in App. S1.

Variable importance analysis

To investigate the relative importance of predictor variables in the models, we built a final set of models for each plant species. The models contained the topo-climatic predictors and those edaphic variables that caused AUC amelioration for most species in the first part of the analysis. For each species, the importance of each variable in the models was assessed in BIOMOD by randomising each variable individually and then recalibrating the model with the randomised variable while keeping the other variables unchanged. The results of the model containing the randomised variable were then correlated with those of the original models. Finally, the importance of the variable was calculated as one minus the correlation; higher values indicate predictors that are more important for the model (Thuiller et al., 2009). This analysis was repeated five times for each modelling technique and the resulting variable importance values were averaged. To investigate changes in predictive power, the variable importance values (averaged

among the four modelling techniques) were displayed for all species as a function of the ecological characteristics and functional traits mentioned above.

Community composition prediction

For each model the predicted probability at each sampling plot was transformed into binary presences or absences by maximising the sensitivity and specificity of the prediction threshold. By stacking the binary prediction for each species, we could predict a community composition for each plot. This stacking of model outputs was performed for each of the nine sets of predictors (topo-climatic only, topo-climatic plus each of the above cited edaphic variables, topo-climatic plus the edaphic variable chosen for the variable importance analysis) and each of the four modelling techniques. We obtained 36 different predicted communities for each plot. Those predicted communities were evaluated by comparison with the observed dataset by computing the Sørensen index (Sørensen, 1948), which estimates the similarity between the predicted and the observed communities, as follow:

$$Sørensen\ index\ = \frac{2a}{2a + b + c}$$

where *a* is the number of species that are observed as well as predicted as present, *b* the number of species observed as present but predicted as absent, *c* is the number of species observed as absent but predicted as present. A Sørensen index of 0 means that there is no agreement between the predicted and observed community while a value of 1 indicate a perfect match between both communities. The correlation between observed and predicted species richness as well as the mean absolute error of species richness were also computed. Species richness is a simple and widely used index of biodiversity that can be modelled by stacking individual species distribution models. However, this method has been demonstrated to over-estimate species richness (Dubuis et al., 2011). We aim at investigating if the addition of edaphic variables in the models can allow reducing this bias. The three evaluation metrics were averaged for the four modelling techniques for each set of predictors. The results obtained were compared between models with and without the edaphic predictor to assess the relevance of the respective edaphic variable to predict plant communities.

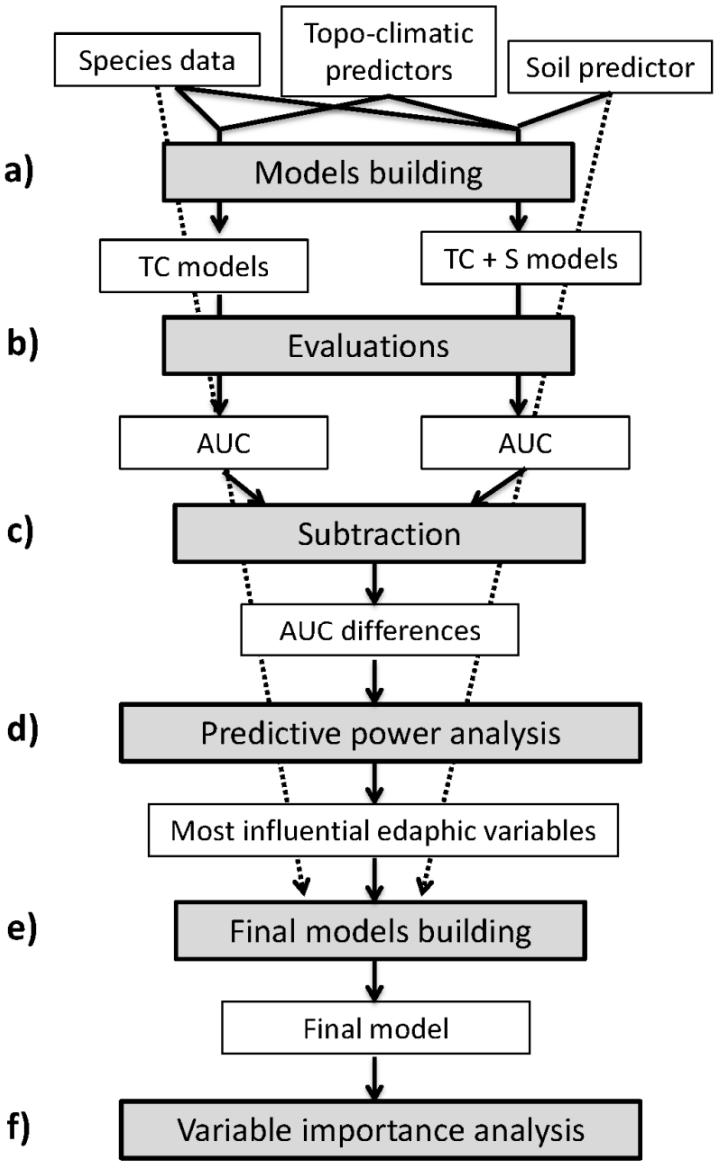


Fig. 2. Flowchart summarizing the sequence of the analysis. For each species four modelling techniques were used to build sets of eight models (a), first with topo-climatic predictors only (TC models), then with topo-climatic and each of the seven edaphic variables (TC + S models). Models were evaluated with AUC (b), and differences in AUC were computed between TC models and each TC + S model (c). Differences in AUC (the mean across the four modelling techniques) were analysed to assess their respective influence in the models across the species (as a function of their ecological characteristics and functional traits) (d). A final model set was built for each species with topo-climatic predictors and the edaphic predictors that most ameliorated the first models (e). A variable importance analysis was made on these more parsimonious models (f). The importance of the edaphic variables conserved in the final models was related to species ecological characteristics and species traits to characterise species for which the retained edaphic variables were important in the models. We finally predict vegetation communities by stacking SDM results and evaluate the resulting prediction with the Sørensen index and by comparing the predicted species richness with the observed richness (not shown in the flowchart)

RESULTS

Model results

Models containing topo-climatic predictors had AUC values ranging from 0.52 to 0.93 (mean = 0.76, App. S1) and TSS values from 0.16 to 0.84 (mean = 0.51, App. S2). These values represented model performance levels ranging from poor to very good. Because the AUC and TSS values are consistent (correlation from 0.93 to 0.95, App. S6), we will focus the following results and discussion sections on the AUC results. The AUC were comparable among the four modelling techniques in the eight models sets, with no technique giving higher or lower evaluation values (App. S7). The correlations computed between the four modelling techniques, for the changes due to the addition of edaphic predictors were all significant and ranged from 0.39 and 0.89 (App. S4). Therefore we consider the average change in AUC among the four modelling technique in the following presentation of results.

Soil pH was the edaphic variable that best improved the models; it positively influenced the models for 47% of the species, with an average AUC increase of 0.05 for the improved models and 0.02, on average, for all models (Table 1).

Table 1. A summary of the changes induced in the AUC upon including edaphic variables in the topo-climatic models. The table lists the mean change for all models, the percentage of species for which the AUC increase is higher than 0.02 and the mean increase for these species, the percentage of species for which the AUC decrease is higher than 0.02 and the mean decrease for these species, and the percentage of species for which the AUC is not strongly influenced by the addition of edaphic variables in the models.

		mean AUC change	% of species with AUC increase of more than 0.02	mean AUC increase	% of species with AUC decrease of more than -0.02	mean AUC decrease	% of species with AUC change between -0.02 and +0.02
AUC	C:N	0.003	29.6	0.035	25.2	-0.036	45.2
	N	0.005	25.2	0.042	21.7	-0.030	53.0
	pH	0.018	47.0	0.049	20.9	-0.032	32.2
	P	-0.002	21.7	0.035	29.6	-0.032	48.7
	Cl	0.002	26.1	0.032	24.3	-0.033	49.6
	Sa	-0.001	21.7	0.033	24.3	-0.035	53.9
	Si	0.000	21.7	0.038	22.6	-0.037	55.7

The species for which models were most ameliorated by the addition of pH were *Vaccinium myrtillus* (an average AUC increase of 0.18 ± 0.02; App. S1), *Nardus stricta* (average AUC increase of 0.12 ± 0.01), *Alchemilla vulgaris, Linum catharticum* and *Hieracium lactucella* (an average AUC increase of 0.1 ± 0.03 for the three species). Total nitrogen content contributed the second most important AUC increase (0.04, on average, for the improved species and 0.005 for all species) and improved the models for 25% of the species. Total nitrogen was important in the models for *Prunella grandiflora* (an average AUC increase of 0.1 ± 0.02), *Veronica serpyllifolia* (an average AUC increase of 0.09 ± 0.01), *Deschampsia cespitosa* (an average AUC increase of 0.09 ± 0.04), and *Knautia dipsacifolia* (an average AUC increase of 0.08 ± 0.01). The carbon-to-nitrogen ratio and clay content positively changed the model for more than half of the species, but only slightly (0.003 and 0.002 on average, respectively). The phosphorus, silt and sand contents changed the AUC negatively for at least half of the models.

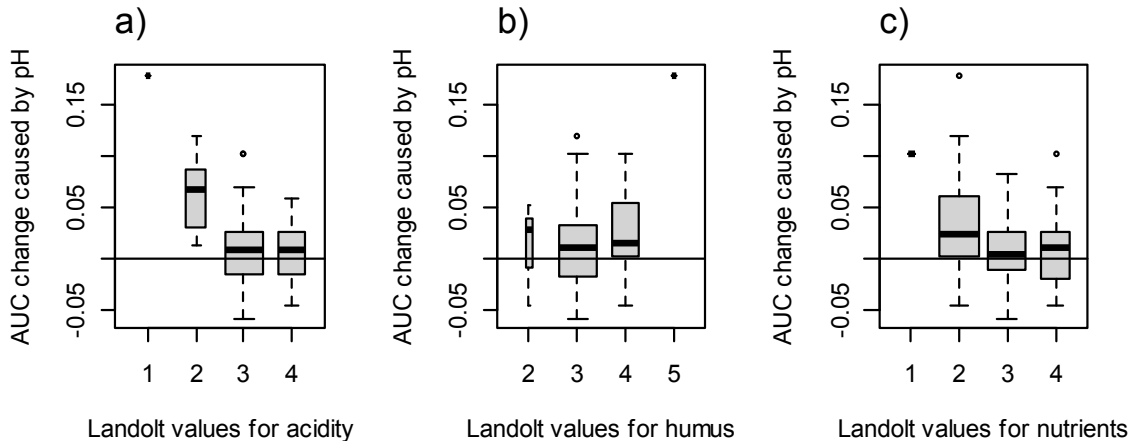


Fig. 3. The change in the AUC upon the addition of edaphic predictors in the models, compared to topo-climatic models, for each species (mean of the four modelling techniques). The plants are classified according to their preferred soil conditions regarding (a) soil acidity, (b) soil humus content and (c) soil nutrient content, as in Landolt et al. (2010). The width of the box-plots is proportional to the number of species showing the indicator value of interest (there was only one species in the categories 1 of acidity and nutrient and the category 5 of humus). The middle line is the median, and the outer limits of the box-plots are the 1st and 4th quartiles.

Species characteristics and predictive power

The predictive power of the models for the small number of species growing preferentially on acidic soil, as determined by their ecological indicator values (Landolt et al., 2010), was improved by adding soil pH to the model (acidity values 1 and 2, Fig. 3a). The models of species with preferences for high humus content and low soil nutrient values displayed a trend towards model amelioration, but these results have to be carefully considered due to the small number of species that present these characteristics (Fig. 3b, c). The AUC changes caused by the other edaphic predictors did not show any trends that could be related to ecological indicator values (App. S8)

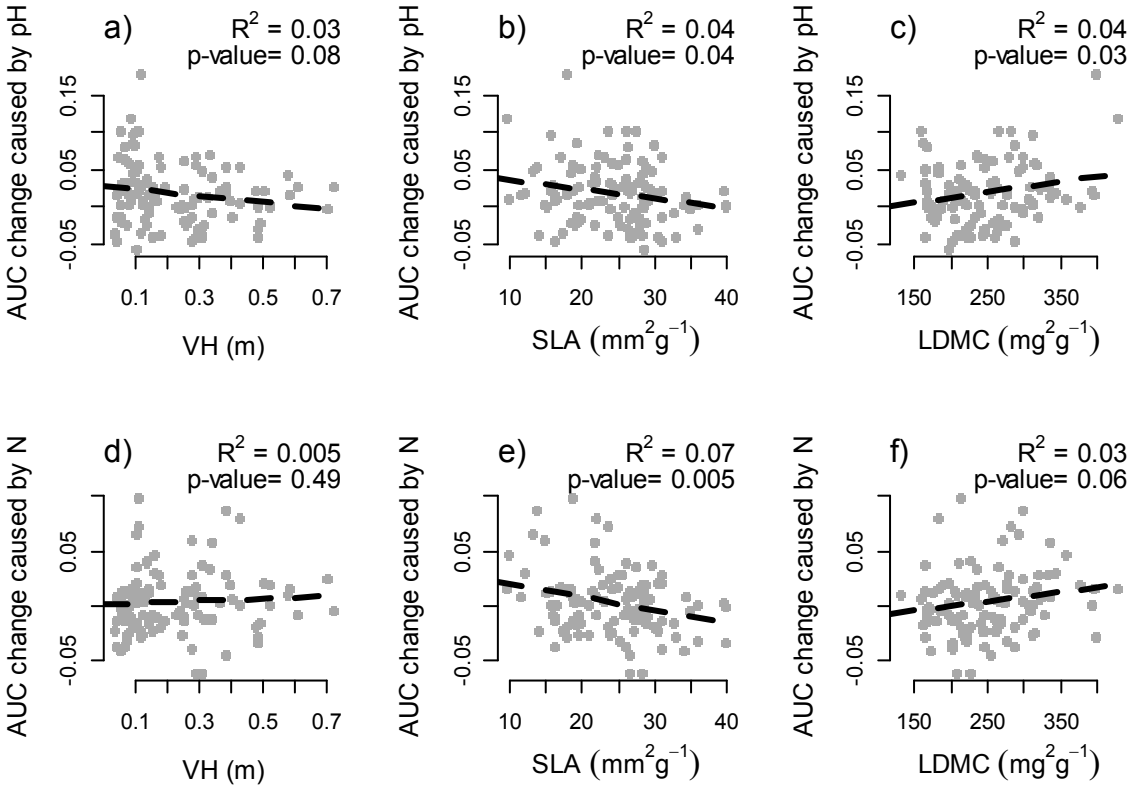


Fig. 4. AUC changes as a function of plant functional traits. The first line shows the change in the AUC, induced by adding pH as a predictor in the species distribution models (mean of the four modelling techniques), as a function of the species (a) VH, (b) SLA and (c) LDMC. The second line shows the change in AUC induced by adding nitrogen as a function of the species (d) VH, (e) SLA and (f) LDMC. The dashed line is the linear regression.

The analysis of the functional traits revealed a significant negative relationship between the average AUC change for each species caused by the addition of nitrogen content as a predictor in the models and its SLA (F = 8.32, degrees of freedom = 103, p-value = 0.005; Fig. 4e). The addition of pH to the models also negatively influenced species with high SLA (F = 4.18, degrees of freedom= 103, p-value = 0.043; Fig. 4b) but positively influenced species with high LDMC (F = 4.73, degrees of freedom = 103, p-value = 0.032; Fig. 4c). No other significant relationships could be demonstrated (App. S9).

Species characteristics and the importance of variables

To build the final model, we used pH and nitrogen content to supplement the topo-climatic variables because pH and nitrogen content were the most influential edaphic variables in our dataset. Comparing all variables based on their importance to the models revealed degree-days and soil pH as the two most important predictors among all species, followed by slope, moisture index and solar radiation. On average among all species and all modelling techniques, nitrogen was the second least important variable (Table 2).

Table 2. A summary of variable importance, as calculated by BIOMOD for each modelling technique, and their means; the larger the value, the more important the variable is in the model. For this analysis only the two edaphic variables that most improved the AUC in previous analysis (pH and N) were kept in addition to topo-climatic predictors.

	GAM	GBM	GLM	RF	Mean
Degree-days	0.494	0.348	0.534	0.258	0.408
pH	0.198	0.145	0.195	0.117	0.164
Slope	0.163	0.137	0.168	0.108	0.144
Moisture	0.129	0.108	0.162	0.108	0.127
Radiation	0.096	0.073	0.106	0.068	0.086
Nitrogen	0.109	0.070	0.100	0.063	0.085
Topographic position	0.047	0.042	0.066	0.056	0.053

Sorting the variable importance results for each species according to the ecological indicator values for soil acidity showed that soil pH was more important in models of species that tolerate high soil acidity (acidity value 1 and 2, Fig. 5a) and high humus content (humus values 4 and 5, Fig. 5b). However, no evidence could be found for the importance of soil nitrogen content (Fig. 5d, e, f).

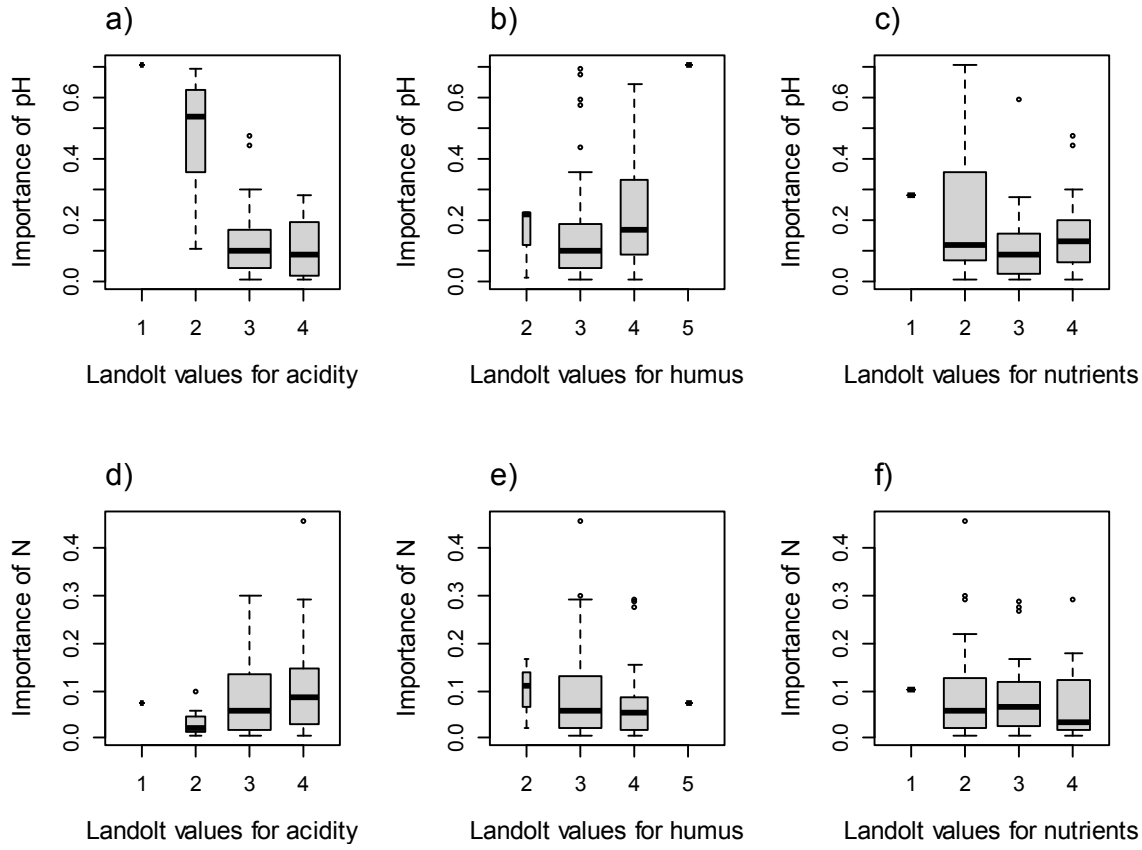


Fig. 5. Importance of edaphic variables as a function of species ecological preferences. The importance (mean of the four modelling techniques) of pH (first row) and N (second row) used as predictors in the models for each species, for the species classified according to their preference for soil acidity (a and d), humus (b and e) and nutrient content (c and f) according to ecological indicator values (Landolt et al. 2010). The width of the box-plots is proportional to the number of species showing the indicator value of interest (there was only one species in the categories 1 of acidity and nutrient and the category 5 of humus). The middle line is the median, and the outer limits of the box-plots are the 1st and 4th quartiles.

Considering the variable importance as a function of the three continuous functional traits we showed that pH importance in the models was related to species VH; pH was significantly more important for species with smaller stature ($F = 4.14$, degrees of freedom = 103, p-value = 0.044, Fig. 6a). N importance showed no significant relationship with VH (Fig. 6d) and no significant relationship was detected for the importance of either edaphic variable as a function of SLA or LDMC (Fig. 6b, c, e and f).

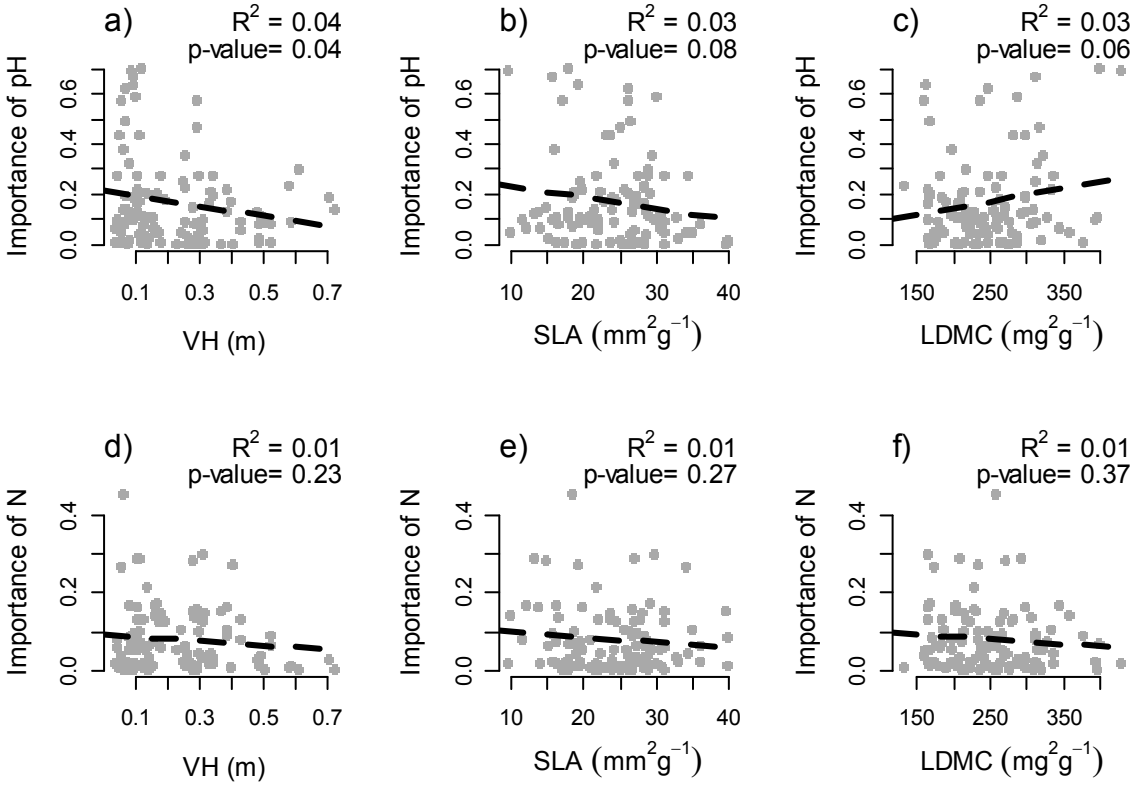


Fig. 6. Importance of edaphic variables as a function of species functional traits. The first plot row shows the importance of pH as a predictor in species distribution models (the mean of the four modelling techniques) as a function of the species (a) VH, (b) SLA and (c) LDMC. The second line shows the importance of nitrogen as a function of the species (d) VH, (e) SLA and (f) LDMC. The dashed line is the linear regression.

Community composition predictions

Among the seven edaphic variables used in the species models, pH was the predictor that was most effective in improving the community predictions. It allowed reductions in the mean absolute error of species richness prediction and increase the correlation between the observed and predicted richness as well as the Sørensen index. Nitrogen produced the second greatest improvement in the prediction of community composition. However, the best results for community prediction were obtained by stacking models containing pH and N as well as topo-climatic predictors. All results are presented in Table 3.

Table 3. Results of the community prediction evaluation. Models for each species were stacked in order to reconstruct a predicted community at each plot. These predictions were evaluated with correlation between observed and predicted species richness (SR), mean absolute error of SR and the Sørensen index. Results presented here for each predictors combinations are averages on modelling techniques and plots.

	Correlation observed-predicted SR	Mean absolute error SR	Mean Sørensen index
TC	0.794	9.838	0.705
TC-C:N	0.828	9.280	0.718
TC-N	0.828	9.093	0.720
TC-pH	0.837	8.736	0.730
TC-P	0.805	9.329	0.716
TC-Cl	0.820	9.300	0.716
TC-Sa	0.811	9.248	0.717
TC-Si	0.827	9.253	0.716
TC-N-pH	0.859	8.164	0.741

DISCUSSION

Our study showed that among the seven edaphic variables studied, pH was the most important predictor of plant species distribution and community composition in our mountain landscape if considered in conjunction with topo-climatic predictors. We also showed that the importance of pH varied across species according to their ecological preferences and morphologies.

How important are soil predictors?

Soil acidity was the most important variable among the soil predictors and was second overall, after degree-days. The importance of degree-days to predict species distributions was expected, given the very large elevation gradient in the study area. The importance of soil acidity was also expected because it affects the plants through nutrient availability and the release of toxic elements; to date, soil acidity has rarely been examined in a modelling study. Peppler-Lisbach & Schröder (2004) previously showed the importance of soil acidity for predicting the distribution of species belonging to communities dominated by *Nardus stricta.* Coudun et al. (2006) and Coudun & Gegout (2007) similarly observed the importance of this factor for *Acer campestre* and *Vaccinium myrtillus*.

The N and P content values were poorer predictors of plant species distribution in our models. This outcome may be due to the fact that these variables, as measured in this study, may not correspond directly to the amounts of nitrogen and phosphorus that are actually available to the plants. For example, high nitrogen content can be associated with a large amount of organic matter and would consequently not be directly available to the plants. Data on available nitrogen and phosphorus would be more interesting for the present type of study, but these measures are also more time consuming, costly and difficult to obtain due to their high variability throughout the year (Cain et al., 1999). Coudun & Gegout (2007) and Pinto & Gegout (2005) used the carbon-to-nitrogen ratio as a proxy for available nitrogen and showed that it is important for predicting the distribution of *Vaccinium myrtillus* as well as the dominant tree species in the Vosges Mountains. We were not able to demonstrate the same effect with our data on non-woody mountain plants.

The texture data (sand, silt and clay content) were no good predictors of our species' distributions, contrary to what would be expected based on several previous studies (Barrett, 2006; Peper et al., 2010; Kamrani et al., 2011). These predictors may have lacked influence in the models because our study area, although reasonably large (700

km^2) and encompassing a wide environmental gradient (400 to 3200 m elevation), does not span the entire soil texture gradient, and therefore lacks plots displaying extreme values of silt, sand or clay content (App. S6).

The results of the community composition prediction analysis confirm the importance of pH for obtaining reliable community composition estimates. Nevertheless, community predictions obtained by stacking single-species distribution models still overestimate the number of species (Dubuis et al. 2011), even if the addition of pH or pH plus nitrogen allows this overestimation to be reduced (from approximately ten to eight species). The similarity between observed and predicted community composition (measured with the Sørensen index) is also slightly improved by adding pH and pH plus N.

One limitation of our study is that we used variables from interpolated maps (the climatic predictors) and variables directly measured in the plot (the soils predictors) in the same models. This approach may have influenced our results in the sense that interpolated variables contain uncertainty (Foster et al., 2012) and may consequently be expected to perform less well as predictors in the models while directly measured variables are more accurate and should be better predictors. This consideration notwithstanding, the best predictor among all those considered remained a climatic one.

Is the importance of edaphic variables for species distribution related to species ecology?

As a predictor of the species distribution in the models, soil acidity was not equally important across all species. Instead, it was related to species characteristics. We related the differences in predictive power between the models with and without edaphic data to the ecological indicator values for soil pH, and we found the highest increase for species that only tolerate acidic soil. We could have expected both acidophilic and basophilic species to respond to the inclusion of soil pH in the models. However its importance for basophilic species did not differ from its importance for neutrophilic species. This finding can be explained by the observation that the parent soil in our study area is calcareous. Thus, acidophilic species can only grow where the soil is deep enough to isolate the roots from the effect of the calcareous bedrock because an excess of calcium would be toxic (Chytry et al., 2003; Gobat et al., 2004). Accounting for soil acidity allows a finer discrimination of these particular cases. Including soil pH as a predictor in the model therefore acts as a filter that better allows

predictions of a species' absence in areas with a suitable climate but in which the species cannot grow due to the basic nature of the soil. We can expect this phenomenon to be inverted in areas with dominant acidic soils, where species less tolerant to acidity will be unable to grow except in areas that contain calcareous inclusions.

It is possible that similar effects could have been observed with the other ecological gradients if an extended data set had been used. For instance, a slight trend across species with ecological values for humus and nutrients was seen in the models improved by pH (Fig. 3b and c), but too few species were available for the analysis to produce significant results. Species in extreme conditions (indicator values of 1 or 5) are rare in our data set, and often have too few occurrences to allow building models of acceptable quality. The majority of species in the analysis had values of 3 or 4, which represents a reduced gradient. Supplementary sampling of areas with more extreme (but also less frequent) site conditions (very high pH, humus or nutrient content) would allow more comprehensive testing.

Is the importance of edaphic variables for species distribution related to species functional traits?

We showed that the AUC increase caused by adding edaphic variables to the models (either pH or nitrogen) was more important for species with a low SLA. pH also significantly improved the models for species with high LDMC. Finally, the importance of pH in SDMs could be demonstrated for a portion of the small-stature species. Although the relationships between changes in AUC and plant functional traits are weak and marginally significant, they can nonetheless be interpreted. These functional attributes (low SLA and VH, high LDMC) correspond to species that grow slowly and sequester nutrients. In our study area, such species are characteristic of nutrient-poor, acidic soils (Rusch et al., 2009; Pontes et al., 2010). These soils display little microbial activity and, consequently, slow litter decomposition and low nutrient availability. Species that require large amounts of nutrients cannot grow under such conditions, whereas those that possess resource-sequestering strategies (i.e., long leaf life span, high dry matter content in the leaves) have an advantage (Aerts & Chapin, 2000). *Vaccinium myrtillus* and *Nardus stricta* are typical examples of such species. Their LDMCs are among the highest, as are their AUC improvements if pH is included in their models.

Conclusion and perspectives

Overall, our study of a large pool of mountain species showed that even if only a minority of edaphic factors significantly improved the models for the species considered, soil acidity should not be overlooked in modelling species and community distributions. SDMs used in conservation planning or for assessing the distribution of rare species could benefit from the inclusion of soil pH and allow for a finer discrimination of favourable sites than SDMs based on climate variables alone, at least, as our results showed, for acidophilic species. The first step to achieve such goals is to build spatially explicit layer for soil characteristics. This could be achieved first by sampling soil at the same time than the vegetation, and then by applying a modelling procedure as proposed in McBratney et al. (2003) or Shi et al. (2009). However, such modelling even if giving good results in the mentioned studies still require to be tested and adapted for areas with a more complex topography. Using better topo-climatic data based on more recent and finer scale measurements would also certainly help to improve the SDM results.

In the context of perspectives for future research, our results also have implications for global change projections. SDMs have been extensively used for projecting future plant species distributions according to several climate or land-use change scenarios (Dirnböck et al., 2003; Thuiller et al., 2005; Engler et al., 2011; Vicente et al., 2011). However, it is known that soil properties are not stable over time and are subject to on-going changes as a result of N deposition that leads to, among other transformations, acidification and eutrophication (Bouwman et al., 2002; Horswill et al., 2008; Bobbink et al., 2010). This process affects the vegetation in the long term (Sala et al., 2000). In future global change modelling studies, it would be interesting to consider incorporating soil predictors in SDMs. Several soil acidification models have been developed (Reinds et al., 2008; Posch & Reinds, 2009), and it would be worthwhile to add soil change scenarios to global change projections, even at coarse resolutions.

ACKNOWLEDGEMENTS

We are grateful to all who helped with the data collection, especially S. Godat, C. Purro, J.-N. Pradervand, and V. Rion. We thank E. Verecchia, J.-M. Gobat and T. Adatte for advice on soil analysis, as well as B. Bomou and T. Monnier for their help in the lab. We also thank G. Litsios for helpful discussions and three anonymous referees for comments that greatly improved this manuscript. This study received support from the Swiss National Fund for Research (SNF grant Nr. 31003A-125145, BIOASSEMBLE project) and from the European Commission (ECOCHANGE Project).

REFERENCES

Aerts, R. & Chapin, F.S. (2000) The mineral nutrition of wild plants revisited: A re-evaluation of processes and patterns. *Advances in Ecological Research, Vol 30* (ed. by A.H. Fitter and D.G. Raffaelli), pp. 1-67. Elsevier Academic Press Inc, San Diego.

Aeschimann, D. & Heitz, C. (1996) *Index synonymique de la flore de Suisse et territoires limitrophes.* CRSF, Geneva, CH.

Allouche, O., Tsoar, A. & 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.

Alvarez, N., Thiel-Egenter, C., Tribsch, A., Holderegger, R., Manel, S., Schonswetter, P., Taberlet, P., Brodbeck, S., Gaudeul, M., Gielly, L., Kupfer, P., Mansion, G., Negrini, R., Paun, O., Pellecchia, M., Rioux, D., Schupfer, F., Van Loo, M., Winkler, M., Gugerli, F. & IntraBioDiv, C. (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632-640.

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

Baize, D. (2000) *Guide des analyses en pédologie*. INRA, Paris.

Barrett, G. (2006) Vegetation communities on the shores of a salt lake in semi-arid Western Australia. *Journal of Arid Environments*, **67**, 77-89.

Bobbink, R., Hicks, K., Galloway, J., Spranger, T., Alkemade, R., Ashmore, M., Bustamante, M., Cinderby, S., Davidson, E., Dentener, F., Emmett, B., Erisman, J.W., Fenn, M., Gilliam, F., Nordin, A., Pardo, L. & De Vries, W. (2010) Global assessment of nitrogen deposition effects on terrestrial plant diversity: a synthesis. *Ecological Applications*, **20**, 30-59.

Bouët, M. (1985) *Climat et météorologie de la Suisse romande*. Payot, Lausanne.

Bouma, T.J. & Bryla, D.R. (2000) On the assessment of root and soil respiration for soils of different textures: interactions with soil moisture contents and soil CO2 concentrations. *Plant and Soil*, **227**, 215-221.

Bouwman, A.F., Van Vuuren, D.P., Derwent, R.G. & Posch, M. (2002) A global analysis of acidification and eutrophication of terrestrial ecosystems. *Water Air and Soil Pollution*, **141**, 349-382.

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

Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145-1157.

Cain, M.L., Subler, S., Evans, J.P. & Fortin, M.J. (1999) Sampling spatial and temporal variation in soil nitrogen availability. *Oecologia*, **118**, 397-404.

Chytry, M., Tichy, L. & Rolecek, J. (2003) Local and regional patterns of species richness in central European vegetation types along the pH/calcium gradient. *Folia Geobotanica*, **38**, 429-442.

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

Coudun, C. & Gegout, J.C. (2007) Quantitative prediction of the distribution and abundance of Vaccinium myrtillus with climatic and edaphic factors. *Journal of Vegetation Science*, **18**, 517-524.

Coudun, C., Gegout, J.C., Piedallu, C. & Rameau, J.C. (2006) Soil nutritional factors improve models of plant species distribution: an illustration with Acer campestre (L.) in France. *Journal of Biogeography*, **33**, 1750-1763.

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.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122-1131.

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. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

Engler, R., Randin, C.F., Vittoz, P., Czaka, T., Beniston, M., Zimmermann, N.E. & Guisan, A. (2009) Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? *Ecography*, **32**, 34-45.

Engler, R., Randin, C.F., Thuiller, W., Dullinger, S., Zimmermann, N.E., Araujo, M.B., Pearman, P.B., Le Lay, G., Piedallu, C., Albert, C.H., Choler, P., Coldea, G., De Lamo, X., Dirnbock, T., Gegout, J.C., Gomez-Garcia, D., Grytnes, J.A., Heegaard, E., Hoistad, F., Nogues-Bravo, D., Normand, S., Puscas, M., Sebastia, M.T., Stanisci, A., Theurillat, J.P., Trivedi, M.R., Vittoz, P. & Guisan, A. (2011) 21st century climate change threatens mountain flora unequally across Europe. *Global Change Biology*, **17**, 2330-2341.

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.

Foster, S.D., Shimadzu, H. & Darnell, R. (2012) Uncertainty in spatially predicted covariates: is it ignorable? *Applied Statistics*, **61**

Friedman, J.H., Hastie, T. & Tibshirani, R. (2000) Additive Logistic Regression: a Statistical View of Boosting. *Annals of Statistics*, **28**, 337-374.

Gobat, J.-M., Aragno, M. & Matthey, W. (2004) *The living soil. Fundamentals of soil science and soil biology*. Science Publishers Inc., Enfield (USA).

Grime, J.P. (1977) Evidence for existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *American Naturalist*, **111**, 1169-1194.

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

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. (2007) What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, **77**, 615-630.

Hastie, T. & Tibshirani, R. (1990) *Generalized Additive Models*. Chapman and Hall, London.

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

Horswill, P., O'Sullivan, O., Phoenix, G.K., Lee, J.A. & Leake, J.R. (2008) Base cation depletion, eutrophication and acidification of species-rich grasslands in response to long-term simulated nitrogen deposition. *Environmental Pollution*, **155**, 336-349.

Kamrani, A., Jalili, A., Naqinezhad, A., Attar, F., Maassoumi, A.A. & Shaw, S.C. (2011) Relationships between environmental variables and vegetation across mountain wetland sites, N. Iran. *Biologia*, **66**, 76-87.

Körner, C. (1999) *Alpine plant life*. Springer, Heidelberg.

Landolt, E., Bäumler, B., Erhardt, A., Hegg, O., Klötzli, F., Lämmler, W., Nobis, M., Rudmann-Maurer, K., Schweingruber, F.H., Theurillat, J.-P., Urmi, E., Vust, M. & Wohlgemuth, T. (2010) *Flora Indicativa. Ecological indicator values and biological attributes of the Flora of Switzerland and the Alps.* Haupt Verlag, Bern.

Marage, D. & Gegout, J.C. (2009) Importance of soil nutrients in the distribution of forest communities on a large geographical scale. *Global Ecology and Biogeography*, **18**, 88-97.

Martre, P., North, G.B., Bobich, E.G. & Nobel, P.S. (2002) Root deployment and shoot growth for two desert species in response to soil rockiness. *American Journal of Botany*, **89**, 1933-1939.

McBratney, A.B., Santos, M.L.M. & Minasny, B. (2003) On digital soil mapping. *Geoderma*, **117**, 3-52.

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

Pellissier, L., Pottier, J., Vittoz, P., Dubuis, A. & Guisan, A. (2010a) Spatial pattern of floral morphology: possible insight into the effects of pollinators on plant distributions. *Oikos*, **119**, 1805-1813.

Pellissier, L., Brathen, K.A., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A., Yoccoz, N.G., Alm, T., Zimmermann, N.E. & Guisan, A. (2010b) Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, **33**, 1004-1014.

Peper, J., Pietzsch, D. & Manthey, M. (2010) Semi-arid rangeland vegetation of the Greater Caucasus foothills in Azerbaijan and its driving environmental conditions. *Phytocoenologia*, **40**, 73-90.

Peppler-Lisbach, C. & Schroder, B. (2004) Predicting the species composition of Nardus stricta communities by logistic regression modelling. *Journal of Vegetation Science*, **15**, 623-634.

Peter, M., Edwards, P.J., Jeanneret, P., Kampmann, D. & Luscher, A. (2008) Changes over three decades in the floristic composition of fertile permanent grasslands in the Swiss Alps. *Agriculture Ecosystems & Environment*, **125**, 204-212.

Pinto, P.E. & Gegout, J.C. (2005) Assessing the nutritional and climatic response of temperate tree species in the Vosges Mountains. *Annals of Forest Science*, **62**, 761-770.

Pontes, L.D., Louault, F., Carrere, P., Maire, V., Andueza, D. & Soussana, J.F. (2010) The role of plant traits and their plasticity in the response of pasture grasses to nutrients and cutting frequency. *Annals of Botany*, **105**, 957-965.

Posch, M. & Reinds, G.J. (2009) A very simple dynamic soil acidification model for scenario analyses and target load calculations. *Environmental Modelling & Software*, **24**, 329-340.

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C., Vittoz, P. & Guisan, A. (2012) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology & Biogeography*, **In press**

Randin, C.F., Vuissoz, G., Liston, G.E., Vittoz, P. & Guisan, A. (2009a) Introduction of Snow and Geomorphic Disturbance Variables into Predictive Models of Alpine Plant Distribution in the Western Swiss Alps. *Arctic Antarctic and Alpine Research*, **41**, 347-361.

Randin, C.F., Jaccard, H., Vittoz, P., Yoccoz, N.G. & Guisan, A. (2009b) Land use improves spatial predictions of mountain plant abundance but not presence-absence. *Journal of Vegetation Science*, **20**, 996-1008.

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.

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

Reinds, G.J., van Oijen, M., Heuvelink, G.B.M. & Kros, H. (2008) Bayesian calibration of the VSD soil acidification model using European forest monitoring data. *Geoderma*, **146**, 475-488.

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

Rusch, G.M., Skarpe, C. & Halley, D.J. (2009) Plant traits link hypothesis about resource-use and response to herbivory. *Basic and Applied Ecology*, **10**, 466-474.

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. & Wall, D.H. (2000) Biodiversity - Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770-1774.

Shi, W.J., Liu, J.Y., Du, Z.P., Song, Y.J., Chen, C.F. & Yue, T.X. (2009) Surface modelling of soil pH. *Geoderma*, **150**, 113-119.

Sørensen, T.A. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter*, **5**, 1–34.

Spaltenstein, H. (1984) *Pédogénèse sur calcaire dur dans les hautes alpes calcaires.* Ecole Polytechnique fédérale de Lausanne, Lausanne.

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285-1293.

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

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.

Vicente, J.V.J., Randin, C.F., Goncalves, J., Metzger, M.J., Lomba, A., Honrado, J. & Guisan, A. (2011) Where will conflicts between alien and rare species occur after climate and land-use change? A test with a novel combined modelling approach. *Biological Invasions*, **13**, 1209-1227.

Zechmeister, H.G., Schmitzberger, I., Steurer, B., Peterseil, J. & Wrbka, T. (2003) The influence of land-use practices and economics on plant species richness in meadows. *Biological Conservation*, **114**, 165-177.

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

Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S. & Blackard, J.A. (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology*, **44**, 1057-1067.

SUPPLEMENTARY INFORMATIONS

Table S1. For each species, this table shows the functional traits and ecological indicator values used in the analysis as well as the change in the AUC caused by the inclusion of edaphic variables in the models. The results are given for each species and each edaphic variable along with the means of the four modelling techniques. H = vegetative height of the plant (in meter), SLA = specific leaf area (in mm^2*g^{-1}), LDMC = leaf dry matter content (in $mg*g^{-1}$); H = humus, Nu = nutrient, A = acidity; TC = the model with topo-climatic predictors only, C:N = carbon : nitrogen ratio, N = nitrogen content, pH = soil pH, P = phosphorus content, Cl = clay content, Sa = sand content and Si = silt content.

	Functional traits			Landolt indicator values			AUC TC		Differences in AUC caused by adding the edaphic predictor													
	VH	SLA	LDMC	H	Nu	A	mean	±	TC-C:N	±	TC-N	±	TC-pH	±	TC-P	±	TC-Cl	±	TC-Sa	±	TC-Si	±
Achillea millefolium	0.30	28.01	205.75	3	3	3	0.8	0.01	-0.04	0.02	-0.06	0.02	-0.04	0.02	-0.03	0.03	-0.04	0.02	-0.02	0.02	-0.07	0.01
Agrostis capillaris	0.25	29.22	320.84	3	2	2	0.8	0.01	0.01	0.00	0.01	0.02	0.03	0.01	-0.02	0.00	-0.01	0.02	0	0.01	0.02	0.00
Ajuga reptans	0.10	28.38	194.60	3	3	3	0.7	0.03	-0.05	0.03	0	0.02	-0.06	0.03	-0.03	0.03	0	0.02	-0.01	0.02	-0.04	0.02
Alchemilla conjuncta aggr	0.07	15.56	390.30	3	2	4	0.8	0.01	0	0.01	0.01	0.03	0.02	0.01	0.03	0.01	-0.02	0.02	0.01	0.01	-0.02	0.02
Alchemilla vulgaris aggr	0.11	23.45	280.76	4	4	3	0.7	0.01	0.04	0.03	0.08	0.03	0.11	0.03	0	0.01	-0.01	0.02	0.02	0.03	0.04	0.02
Anthoxanthum odoratum aggr	0.10	29.87	285.80	3	3	2	0.7	0.00	0.01	0.02	-0.02	0.01	0.08	0.00	-0.04	0.02	0.01	0.01	-0.04	0.01	-0.01	0.01
Anthyllis vulneraria sl	0.08	15.95	201.01	3	2	3	0.7	0.02	-0.03	0.02	-0.02	0.02	0.01	0.03	-0.02	0.03	-0.01	0.02	-0.04	0.01	-0.04	0.01
Aposeris foetida	0.08	39.61	167.84	4	3	4	0.7	0.01	0.02	0.02	0	0.01	0	0.00	0.02	0.02	0.01	0.03	0.02	0.02	0.03	0.02
Arnica montana	0.09	19.03	161.00	4	2	2	0.8	0.02	-0.06	0.03	-0.02	0.03	0.09	0.03	0.02	0.01	0.03	0.03	0.02	0.05	-0.01	0.02
Arrhenatherum elatius	0.72	26.50	306.70	3	4	3	0.8	0.01	0.04	0.01	0	0.01	0.03	0.01	-0.01	0.01	0.01	0.04	-0.01	0.02	-0.02	0.04
Aster bellidiastrum	0.05	27.45	170.49	3	2	4	0.7	0.02	-0.01	0.01	0	0.03	-0.02	0.02	-0.02	0.02	0.02	0.00	0.01	0.02	0	0.01
Astrantia major	0.29	26.23	243.29	3	3	4	0.8	0.02	-0.08	0.02	-0.01	0.02	-0.03	0.02	-0.01	0.02	-0.04	0.02	-0.06	0.02	-0.04	0.01
Bartsia alpina	0.13	19.61	264.04	4	3	3	0.8	0.02	-0.01	0.00	0.01	0.00	-0.01	0.01	0.01	0.01	0	0.01	0	0.00	-0.01	0.00
Brachypodium pinnatum aggr	0.52	28.17	374.78	3	3	4	0.8	0.02	0.03	0.02	0.02	0.03	0	0.02	0.01	0.01	0.05	0.02	0.03	0.03	0.04	0.02
Briza media	0.31	27.83	266.39	3	2	3	0.8	0.01	-0.03	0.00	-0.01	0.02	-0.02	0.01	-0.03	0.02	-0.02	0.02	-0.06	0.02	-0.04	0.01
Bromus erectus sstr	0.42	20.62	310.25	3	2	4	0.9	0.02	-0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.00	0	0.01	0.01	0.01	-0.01	0.02
Calamagrostis varia	0.34	21.51	333.04	3	2	4	0.8	0.02	0.04	0.03	0.05	0.01	0.04	0.02	-0.03	0.02	0.02	0.01	-0.02	0.04	-0.01	0.03
Campanula rhomboidalis	NA	NA	NA	3	4	3	0.7	0.01	-0.02	0.04	0	0.02	0.01	0.03	-0.02	0.01	0.01	0.02	0	0.03	0.03	0.01
Campanula rotundifolia	NA	NA	NA	3	2	3	0.7	0.03	-0.03	0.02	0.02	0.02	-0.03	0.02	-0.01	0.00	0.01	0.03	0.02	0.02	-0.01	0.03
Campanula scheuchzeri	0.09	21.95	258.50	NA	3	3	0.9	0.01	0.01	0.00	0.01	0.01	0.03	0.01	0	0.00	0	0.01	0	0.01	-0.01	0.01
Carduus defloratus sstr	0.10	19.56	161.36	2	3	4	0.8	0.01	-0.03	0.01	-0.02	0.02	0.05	0.03	0	0.02	-0.05	0.02	-0.02	0.01	-0.01	0.01
Carex ferruginea	0.24	19.34	310.27	3	3	4	0.8	0.02	-0.01	0.02	-0.02	0.05	0.01	0.00	0	0.01	-0.04	0.01	0.01	0.02	0	0.02
Carex flacca	0.38	25.43	332.74	2	2	4	0.7	0.01	0.03	0.03	0.03	0.03	0.03	0.02	-0.02	0.01	0.03	0.01	0	0.01	-0.02	0.01
Carex montana	NA	NA	NA	3	2	4	0.7	0.02	0.02	0.01	-0.03	0.02	0.01	0.01	0.05	0.01	0.02	0.02	-0.02	0.01	-0.03	0.03
Carex pallescens	0.29	34.22	332.87	4	3	2	0.7	0.00	0	0.01	0.01	0.01	0.01	0.02	0.02	0.00	0.02	0.01	0.01	0.02	0.01	0.01

Carex sempervirens	0.15	9.71	356.78	3	2	3	0.8	0.01	-0.04	0.01	0.04	0.01	0.01	0.01	-0.01	0.01	-0.01	0.00	0.01	0.01	-0.01	0.01
Carlina acaulis ssp caulescens	0.13	15.09	173.21	3	2	3	0.8	0.01	0	0.01	0.01	0.00	0	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0	0.01
Carum carvi	0.25	28.07	256.34	3	3	3	0.8	0.03	-0.04	0.01	-0.01	0.01	0.06	0.01	-0.01	0.02	0.02	0.03	0.04	0.03	0	0.01
Cerastium fontanum ssp vulgare	0.16	28.85	164.13	3	4	3	0.7	0.01	-0.03	0.01	-0.01	0.03	0.01	0.02	0.04	0.02	-0.03	0.01	0.01	0.03	0.01	0.01
Chaerophyllum hirsutum	0.48	35.80	220.29	3	4	3	0.7	0.01	0.01	0.01	-0.03	0.01	-0.03	0.01	0	0.00	-0.05	0.01	0.01	0.03	0.01	0.02
Crepis aurea	0.04	30.85	177.60	4	4	3	0.7	0.01	0	0.01	-0.01	0.01	-0.01	0.02	-0.01	0.02	-0.04	0.02	0	0.01	-0.01	0.01
Crepis pyrenaica	0.27	30.90	201.35	4	4	4	0.8	0.01	-0.02	0.03	-0.03	0.02	-0.04	0.04	-0.04	0.01	-0.01	0.01	-0.01	0.02	-0.01	0.02
Crocus albiflorus	NA	NA	NA	4	4	3	0.8	0.02	-0.02	0.01	0	0.01	0.02	0.02	-0.02	0.01	0	0.01	-0.04	0.01	-0.01	0.01
Cynosurus cristatus	0.14	32.57	238.08	3	3	3	0.9	0.01	-0.02	0.00	-0.02	0.01	-0.01	0.00	-0.03	0.01	-0.02	0.00	-0.05	0.00	-0.05	0.01
Dactylis glomerata sstr	NA	NA	NA	3	4	3	0.8	0.02	0	0.01	0.02	0.00	0.01	0.01	0	0.02	0.02	0.01	-0.02	0.00	0.03	0.01
Daucus carota	0.28	26.58	224.62	3	2	3	0.9	0.01	-0.09	0.05	-0.06	0.02	-0.04	0.01	-0.05	0.04	-0.06	0.02	-0.05	0.04	-0.08	0.04
Deschampsia cespitosa	0.38	13.60	296.14	3	4	3	0.6	0.02	0.06	0.02	0.09	0.04	0.06	0.02	0.05	0.01	0.07	0.02	0.04	0.01	0.04	0.01
Euphorbia cyparissias	0.17	30.79	276.23	3	2	3	0.8	0.03	0.05	0.02	0.03	0.01	0.06	0.03	0.05	0.03	0.02	0.04	0.05	0.03	0.06	0.02
Euphrasia minima	0.07	23.76	215.11	4	2	2	0.9	0.00	0	0.00	-0.03	0.01	0.03	0.00	-0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Festuca pratensis sstr	0.58	24.05	274.15	4	4	3	0.8	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0	0.01	0.01	0.01	0.01	0.01
Festuca rubra aggr	0.15	19.24	327.03	3	3	3	0.7	0.02	0.01	0.03	-0.01	0.02	0.02	0.02	-0.02	0.01	-0.01	0.03	-0.04	0.04	0.01	0.02
Festuca violacea aggr	0.14	17.61	392.86	3	3	2	0.9	0.01	0.02	0.01	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.02	0.02	0.01	0.03	0.01
Galium album	0.33	22.09	234.52	3	4	3	0.7	0.02	-0.01	0.02	0.03	0.03	0.07	0.03	0	0.01	0.02	0.01	0	0.01	0.05	0.02
Galium anisophyllon	0.12	20.60	218.77	3	2	3	0.9	0.01	0	0.01	-0.03	0.03	0	0.01	-0.01	0.02	-0.02	0.01	-0.01	0.01	-0.01	0.00
Galium pumilum	0.13	21.46	225.76	3	2	3	0.6	0.02	0	0.03	0.04	0.02	-0.03	0.03	0	0.04	-0.02	0.03	-0.01	0.04	0.01	0.04
Gentiana campestris sstr	0.10	30.42	195.13	3	2	3	0.8	0.02	0.04	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.03	0.01	-0.01	0.02
Gentiana lutea	0.48	17.84	224.34	3	3	4	0.8	0.04	0	0.01	-0.01	0.01	0.02	0.02	0	0.01	0.04	0.01	-0.03	0.02	0.05	0.02
Gentiana purpurea	0.29	17.01	232.16	3	2	2	0.8	0.02	-0.01	0.01	0.02	0.02	0.07	0.01	-0.03	0.02	-0.03	0.02	0	0.01	0	0.02
Geranium sylvaticum	0.38	26.52	251.57	3	4	3	0.7	0.01	-0.02	0.01	-0.05	0.01	0.02	0.02	0	0.01	-0.01	0.01	-0.01	0.01	0.01	0.01
Geum montanum	0.09	15.51	310.43	3	2	2	0.8	0.02	0.01	0.03	0.02	0.02	0.1	0.03	-0.02	0.01	0.03	0.01	-0.01	0.01	0	0.04
Helianthemum nummularium sl	0.10	13.12	290.63	3	2	4	0.7	0.03	0.03	0.01	0.06	0.02	0.04	0.02	0.03	0.01	0.01	0.01	0.02	0.03	0.07	0.02
Heracleum sphondylium sl	0.33	28.11	204.21	4	4	3	0.7	0.02	-0.03	0.01	-0.01	0.01	0.02	0.02	-0.02	0.01	-0.01	0.02	-0.03	0.01	0.02	0.01
Hieracium lactucella	0.05	26.07	157.70	4	2	2	0.7	0.04	0	0.02	0	0.02	0.1	0.03	-0.03	0.01	0.02	0.02	0.01	0.01	-0.01	0.01
Hieracium murorum aggr	NA	NA	NA	4	3	3	0.6	0.02	0.03	0.04	0.03	0.02	0.03	0.05	0.06	0.03	0.01	0.02	0.07	0.03	0.01	0.02
Hippocrepis comosa	0.05	18.15	255.03	3	2	4	0.8	0.03	-0.04	0.01	0.01	0.03	-0.03	0.02	-0.08	0.03	0.03	0.04	0.05	0.04	0.01	0.02
Holcus lanatus	0.52	35.83	236.98	4	3	3	0.9	0.01	0.02	0.01	0	0.03	0.01	0.03	0.02	0.00	0.01	0.03	0.01	0.01	-0.01	0.01
Homogyne alpina	0.04	11.40	229.40	4	2	3	0.8	0.01	0.03	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	-0.02	0.00	-0.01	0.01
Hypericum maculatum sstr	0.29	24.89	314.34	4	4	3	0.6	0.04	0.02	0.02	0.01	0.03	0.05	0.05	-0.04	0.01	-0.01	0.04	0.01	0.06	-0.03	0.04
Knautia dipsacifolia sstr	0.42	21.86	181.54	3	3	3	0.7	0.02	0.03	0.02	0.08	0.01	0.06	0.04	0.01	0.01	0.1	0.01	0.07	0.01	0.1	0.00
Lathyrus pratensis	0.25	28.45	285.98	3	3	3	0.8	0.01	0	0.01	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.03	-0.01	0.01	0.02	0.02
Leontodon autumnalis	0.05	33.82	170.90	3	3	3	0.6	0.02	0.01	0.02	-0.04	0.03	0	0.02	-0.04	0.01	-0.03	0.04	-0.06	0.05	0.01	0.02
Leontodon helveticus	0.06	26.30	163.97	4	2	2	0.8	0.02	0.03	0.01	0	0.01	0.08	0.02	-0.01	0.00	0	0.01	-0.02	0.01	0.01	0.01

Leontodon hispidus aggr	0.10	27.71	162.63	3	3	3	0.6	0.02	-0.02	0.00	0.01	0.02	0.03	0.02	-0.04	0.01	0.01	0.02	0.02	0.01	-0.02	0.01
Leucanthemum vulgare aggr	0.25	19.30	175.74	3	2	3	0.7	0.01	-0.05	0.01	-0.02	0.02	-0.02	0.01	-0.03	0.01	-0.04	0.00	-0.04	0.01	-0.07	0.01
Ligusticum mutellina	0.06	21.37	253.52	3	3	3	0.8	0.00	0.03	0.01	0.02	0.01	0.03	0.01	0.03	0.01	0	0.01	0	0.01	0.03	0.01
Linum catharticum	0.10	27.27	262.30	3	1	3	0.7	0.03	0.02	0.04	0.03	0.03	0.1	0.03	0	0.03	0.01	0.03	0	0.03	0.04	0.02
Lolium perenne	0.49	28.72	265.08	3	4	3	0.9	0.02	-0.02	0.02	-0.02	0.03	-0.02	0.03	0	0.01	0	0.02	-0.02	0.02	-0.01	0.02
Lotus corniculatus sl	0.11	26.13	230.85	3	3	4	0.6	0.01	0.02	0.01	0.01	0.01	0.02	0.00	0.02	0.00	0.01	0.02	0.04	0.01	0.03	0.01
Luzula multiflora	0.06	25.82	244.60	4	2	2	0.7	0.02	-0.02	0.02	-0.04	0.03	0.06	0.04	0.01	0.01	-0.01	0.02	-0.03	0.03	-0.01	0.02
Nardus stricta	0.08	9.54	428.20	3	2	2	0.8	0.01	-0.01	0.02	0.01	0.01	0.12	0.01	0.01	0.02	0.03	0.02	0.02	0.02	-0.01	0.02
Phleum pratense	0.70	23.81	296.68	3	4	3	0.8	0.02	-0.02	0.01	0.03	0.01	0	0.02	0	0.01	0.01	0.01	-0.01	0.02	0	0.01
Phleum rhaeticum	0.34	27.02	296.67	3	4	3	0.8	0.01	0.03	0.01	-0.01	0.02	0	0.02	0.01	0.01	0	0.02	0	0.01	0.01	0.02
Phyteuma orbiculare	0.17	28.62	244.16	4	2	4	0.7	0.03	-0.02	0.01	0.02	0.01	0.06	0.03	-0.01	0.03	-0.02	0.03	0.01	0.01	-0.05	0.03
Phyteuma spicatum	0.48	39.69	199.53	4	3	3	0.7	0.02	-0.04	0.01	-0.03	0.02	0	0.04	-0.03	0.04	-0.04	0.03	-0.03	0.02	-0.03	0.02
Pimpinella major	0.31	26.71	218.67	3	3	3	0.8	0.02	-0.01	0.01	0.01	0.02	-0.04	0.01	0.03	0.01	0.01	0.02	-0.01	0.01	0	0.02
Plantago alpina	0.06	15.79	194.28	4	2	2	0.9	0.01	-0.01	0.01	-0.01	0.01	0.03	0.01	-0.04	0.02	0	0.01	0	0.01	0	0.01
Plantago atrata sstr	0.08	16.18	208.46	4	3	4	0.8	0.01	0	0.01	-0.01	0.00	0.01	0.01	0.02	0.01	0.01	0.01	-0.01	0.01	-0.02	0.00
Plantago lanceolata	0.23	25.66	165.17	3	3	3	0.9	0.01	0.01	0.01	-0.01	0.01	0	0.01	0	0.01	0	0.01	-0.01	0.00	-0.02	0.00
Plantago media	0.06	18.54	165.96	3	2	4	0.7	0.02	0.01	0.02	-0.01	0.01	0.03	0.01	-0.03	0.01	-0.01	0.01	0.01	0.02	0	0.01
Poa alpina	0.07	25.20	274.54	3	4	3	0.9	0.00	0.02	0.01	-0.03	0.00	-0.02	0.01	0	0.00	0	0.01	-0.04	0.00	-0.03	0.01
Polygala vulgaris	NA	NA	NA	3	2	4	0.7	0.03	-0.02	0.02	0	0.02	0.01	0.02	0.1	0.01	-0.02	0.02	0.02	0.03	0.03	0.02
Polygonum bistorta	0.60	28.64	214.23	4	4	3	0.7	0.02	0.06	0.01	-0.01	0.01	0.02	0.01	0.04	0.02	0	0.01	-0.02	0.01	-0.01	0.02
Polygonum viviparum	0.12	21.86	215.06	4	2	3	0.8	0.01	0	0.01	-0.01	0.00	0.01	0.01	-0.01	0.02	-0.05	0.01	-0.03	0.01	-0.03	0.01
Potentilla aurea	0.04	22.99	295.07	3	2	2	0.8	0.01	0.01	0.02	0.01	0.01	0.07	0.01	-0.02	0.01	0.01	0.02	0.01	0.00	0.01	0.03
Potentilla erecta	0.07	23.93	311.16	4	2	NA	0.8	0.01	0	0.03	-0.02	0.01	0.06	0.01	0.01	0.01	0.02	0.01	0.01	0.03	-0.02	0.02
Primula elatior sstr	0.11	26.68	207.30	4	4	3	0.6	0.03	0	0.01	0.01	0.02	0.02	0.02	-0.01	0.02	-0.01	0.05	-0.01	0.01	-0.05	0.02
Prunella grandiflora	0.11	18.44	211.71	3	2	4	0.8	0.02	-0.01	0.01	0.09	0.02	0.05	0.00	0.05	0.01	0.03	0.01	0.03	0.02	0.02	0.01
Prunella vulgaris	0.13	25.89	209.78	3	3	3	0.8	0.02	0.01	0.01	0.04	0.02	0.04	0.01	0	0.01	-0.02	0.01	-0.01	0.01	-0.01	0.01
Pulsatilla alpina sstr	0.28	16.44	303.90	3	3	4	0.8	0.01	-0.03	0.03	0	0.02	0.01	0.02	0.01	0.02	0.01	0.01	0.05	0.02	0	0.02
Ranunculus acris sl	0.27	26.43	202.15	3	4	3	0.8	0.00	0.04	0.01	0.01	0.01	-0.02	0.01	0	0.01	0	0.01	0	0.00	-0.01	0.01
Ranunculus montanus	0.17	23.44	211.88	3	4	4	0.8	0.01	-0.02	0.01	-0.02	0.02	-0.03	0.02	-0.03	0.01	-0.03	0.01	-0.01	0.01	-0.01	0.01
Ranunculus nemorosus	NA	NA	NA	3	2	3	0.6	0.02	-0.01	0.02	0.06	0.02	0.05	0.02	0.01	0.02	0.02	0.03	0.01	0.02	0.02	0.02
Rhinanthus alectorolophus	0.40	19.05	231.42	4	3	4	0.6	0.02	0.02	0.05	0.01	0.02	-0.02	0.04	0.01	0.03	0.06	0.03	0.02	0.05	0	0.07
Rumex acetosa	0.58	28.05	128.50	4	3	3	0.8	0.01	0.03	0.01	0.01	0.01	0.04	0.01	0	0.01	0	0.01	0.04	0.01	-0.01	0.01
Sanguisorba minor	0.27	24.91	318.88	3	2	4	0.8	0.00	0.02	0.01	0.03	0.01	-0.01	0.01	0.02	0.02	0.02	0.00	0	0.01	0.01	0.00
Scabiosa lucida	0.09	18.34	181.00	3	2	4	0.8	0.01	0	0.01	0.01	0.02	0.04	0.02	0.04	0.01	0.02	0.01	0	0.01	0.02	0.01
Sesleria caerulea	0.16	11.96	342.24	3	2	4	0.9	0.00	0	0.02	0.03	0.00	0.02	0.00	-0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Silene vulgaris sl	0.30	29.37	163.04	3	2	3	0.6	0.03	0.06	0.03	0.04	0.03	0.05	0.01	0.03	0.05	0.03	0.02	-0.01	0.06	0.02	0.04
Soldanella alpina	0.03	13.94	240.22	4	3	3	0.9	0.01	-0.02	0.01	-0.02	0.01	-0.04	0.01	-0.02	0.01	-0.03	0.01	-0.01	0.01	-0.01	0.01

Stellaria graminea	NA	NA	NA	3	3	2	0.8	0.01	0.03	0.02	0.01	0.01	0.03	0.03	-0.03	0.02	0	0.02	-0.03	0.01	-0.03	0.02
Taraxacum officinale	0.23	34.76	184.86	3	4	3	0.8	0.01	-0.02	0.01	-0.01	0.01	-0.01	0.01	-0.02	0.01	-0.02	0.01	0	0.01	-0.02	0.01
Thymus praecox ssp polytrichus	0.04	16.19	285.67	2	2	4	0.7	0.02	-0.07	0.05	-0.04	0.04	-0.05	0.03	-0.06	0.03	-0.07	0.03	-0.05	0.02	-0.02	0.02
Thymus pulegioides sstr	0.03	17.08	261.66	3	2	3	0.8	0.03	0.02	0.02	-0.02	0.01	-0.03	0.01	-0.02	0.02	0.01	0.01	-0.03	0.02	-0.04	0.02
Tragopogon pratensis	0.48	32.71	163.22	3	3	3	0.7	0.03	0	0.03	-0.03	0.01	-0.04	0.03	-0.06	0.02	-0.02	0.02	-0.05	0.01	-0.04	0.02
Trifolium badium	0.11	26.93	243.08	3	3	4	0.7	0.01	-0.02	0.03	-0.02	0.01	-0.03	0.01	0.01	0.05	-0.02	0.02	0.01	0.02	0.02	0.02
Trifolium medium	0.25	18.76	271.47	3	3	3	0.8	0.01	0.01	0.02	0	0.03	-0.02	0.02	0.04	0.02	-0.04	0.02	-0.01	0.03	0.01	0.02
Trifolium pratense sstr	0.19	24.33	259.64	3	3	3	0.7	0.01	-0.01	0.01	-0.02	0.01	0.02	0.02	0.01	0.02	-0.02	0.01	0	0.01	-0.01	0.04
Trifolium repens sstr	0.12	28.52	219.48	3	4	3	0.8	0.01	0.01	0.02	0.01	0.01	0.03	0.01	0.02	0.01	0.02	0.01	-0.02	0.01	0	0.02
Trisetum flavescens	0.50	30.96	274.33	3	4	3	0.7	0.02	0.05	0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.02	0.02	0.03	0.02
Trollius europaeus	0.27	14.73	268.63	4	3	3	0.7	0.02	-0.01	0.01	0.06	0.02	0	0.01	-0.01	0.01	0.03	0.02	0.03	0.01	0.04	0.02
Vaccinium myrtillus	0.11	17.82	396.66	5	2	1	0.7	0.02	0.03	0.02	-0.03	0.03	0.18	0.02	0	0.02	0.03	0.03	0.02	0.02	-0.01	0.03
Veratrum album ssp lobelianum	0.39	22.72	187.34	4	3	3	0.8	0.02	0.02	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.05	0.02	0.01	0.00	0.02	0.01
Veronica chamaedrys	0.15	30.23	257.38	3	4	3	0.8	0.01	0.01	0.01	-0.02	0.01	-0.04	0.02	-0.01	0.01	0	0.01	0.04	0.01	-0.01	0.01
Veronica serpyllifolia sl	NA	NA	NA	3	4	3	0.6	0.03	0.13	0.03	0.09	0.01	-0.01	0.01	0.03	0.04	0.04	0.04	-0.02	0.03	0.06	0.03
Vicia sepium	0.36	39.36	228.36	3	3	3	0.8	0.01	0.02	0.02	0	0.01	0.02	0.01	-0.01	0.01	0.02	0.01	-0.01	0.01	0	0.01

Table S2. The change in the TSS values caused by the addition of edaphic variables in the models. The results are given for each species and each edaphic variable with the mean and standard deviation of the four modelling techniques. TC = model with topo-climatic predictors only, C:N = carbon to nitrogen ratio, N = nitrogen content, pH = soil pH, P = phosphorus content, Cl = clay content, Sa = sand content and Si = silt content.

	TSS TC		Differences in TSS caused by adding the edaphic predictor													
	mean	±	TC-C:N	±	TC-N	±	TC-pH	±	TC-P	±	TC-Cl	±	TC-Sa	±	TC-Si	±
Achillea millefolium	0.47	0.03	-0.03	0.04	-0.08	0.04	-0.05	0.04	-0.04	0.04	-0.05	0.03	-0.01	0.04	-0.08	0.03
Agrostis capillaris	0.49	0.02	0.00	0.02	0.01	0.02	0.05	0.01	-0.01	0.01	-0.05	0.02	0.00	0.01	0.01	0.01
Ajuga reptans	0.46	0.03	-0.07	0.03	0.01	0.03	-0.07	0.03	-0.03	0.04	0.00	0.02	-0.01	0.03	-0.06	0.02
Alchemilla conjuncta aggr	0.50	0.00	0.03	0.04	0.05	0.03	0.05	0.02	0.04	0.03	-0.01	0.04	0.01	0.02	-0.02	0.05
Alchemilla vulgaris aggr	0.36	0.02	0.06	0.05	0.08	0.05	0.15	0.04	-0.02	0.01	-0.01	0.02	0.03	0.05	0.05	0.03
Anthoxanthum odoratum aggr	0.38	0.02	0.00	0.02	-0.02	0.01	0.12	0.02	-0.06	0.02	0.00	0.03	-0.04	0.03	-0.01	0.03
Anthyllis vulneraria sl	0.43	0.05	-0.06	0.02	-0.03	0.03	0.02	0.06	-0.02	0.05	-0.03	0.03	-0.07	0.04	-0.07	0.03
Aposeris foetida	0.41	0.02	0.04	0.04	0.02	0.02	0.00	0.01	0.02	0.03	0.06	0.05	0.03	0.02	0.07	0.04
Arnica montana	0.63	0.02	-0.11	0.06	-0.02	0.03	0.14	0.03	0.02	0.02	0.04	0.02	0.04	0.08	0.00	0.05
Arrhenatherum elatius	0.64	0.04	0.05	0.03	0.01	0.02	0.06	0.03	-0.03	0.04	0.03	0.06	-0.02	0.02	-0.03	0.09
Aster bellidiastrum	0.48	0.02	0.01	0.03	-0.01	0.05	-0.02	0.03	-0.03	0.03	0.06	0.02	0.02	0.01	0.01	0.03
Astrantia major	0.53	0.05	-0.11	0.06	0.00	0.03	-0.05	0.03	0.01	0.04	-0.06	0.04	-0.08	0.04	-0.06	0.04
Bartsia alpina	0.58	0.01	-0.01	0.02	-0.02	0.02	-0.04	0.01	0.00	0.02	0.00	0.03	0.00	0.01	-0.04	0.03
Brachypodium pinnatum aggr	0.60	0.03	0.02	0.01	0.02	0.04	0.00	0.04	0.00	0.03	0.07	0.04	0.06	0.05	0.06	0.03
Briza media	0.50	0.03	-0.06	0.01	0.00	0.02	-0.03	0.01	-0.06	0.03	-0.05	0.02	-0.09	0.03	-0.08	0.02
Bromus erectus sstr	0.70	0.04	0.00	0.06	0.02	0.03	0.02	0.02	0.04	0.06	0.01	0.04	0.05	0.03	0.00	0.05
Calamagrostis varia	0.57	0.04	0.05	0.04	0.10	0.04	0.03	0.07	-0.09	0.01	0.06	0.01	-0.03	0.02	-0.03	0.05
Campanula rhomboidalis	0.42	0.01	0.00	0.04	-0.02	0.01	0.00	0.03	0.00	0.03	0.01	0.02	-0.02	0.04	0.00	0.01
Campanula rotundifolia	0.41	0.05	-0.05	0.03	0.01	0.02	-0.03	0.03	-0.03	0.02	0.01	0.04	0.03	0.04	0.00	0.05
Campanula scheuchzeri	0.67	0.02	0.04	0.01	0.02	0.01	0.03	0.01	0.00	0.01	0.01	0.02	0.00	0.02	-0.02	0.00
Carduus defloratus sstr	0.49	0.03	-0.04	0.02	-0.05	0.04	0.09	0.04	0.00	0.03	-0.07	0.03	-0.02	0.02	0.00	0.02
Carex ferruginea	0.64	0.05	-0.03	0.03	-0.06	0.07	-0.01	0.02	-0.02	0.02	-0.09	0.02	-0.02	0.04	-0.01	0.05
Carex flacca	0.44	0.01	0.05	0.03	0.07	0.04	0.06	0.02	-0.04	0.01	0.05	0.01	0.00	0.03	0.01	0.01
Carex montana	0.48	0.05	0.01	0.01	-0.06	0.04	0.00	0.05	0.04	0.04	0.04	0.05	-0.04	0.03	-0.05	0.07
Carex pallescens	0.45	0.02	0.00	0.01	0.01	0.03	0.02	0.02	0.03	0.02	0.04	0.01	0.01	0.05	0.01	0.02
Carex sempervirens	0.60	0.02	-0.05	0.02	0.08	0.02	0.01	0.04	-0.02	0.02	0.01	0.02	0.02	0.02	0.00	0.03
Carlina acaulis ssp caulescens	0.53	0.02	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.03	0.02	0.02	0.01	0.03	0.01	0.05
Carum carvi	0.54	0.06	-0.06	0.02	-0.03	0.02	0.10	0.05	-0.03	0.02	0.04	0.04	0.05	0.04	0.00	0.03
Cerastium fontanum ssp vulgare	0.37	0.03	-0.05	0.01	0.00	0.05	0.04	0.01	0.03	0.03	-0.05	0.03	0.02	0.04	0.02	0.03

Chaerophyllum hirsutum	0.41	0.03	0.02	0.02	-0.04	0.03	-0.03	0.04	0.01	0.05	-0.06	0.00	0.01	0.02	0.02	0.01
Crepis aurea	0.51	0.01	0.00	0.05	-0.04	0.01	-0.01	0.03	-0.04	0.05	-0.08	0.02	0.01	0.03	-0.04	0.02
Crepis pyrenaica	0.55	0.05	-0.05	0.06	-0.07	0.07	-0.06	0.08	-0.06	0.02	-0.04	0.03	-0.02	0.03	-0.03	0.04
Crocus albiflorus	0.62	0.04	-0.04	0.01	0.03	0.02	0.02	0.05	-0.02	0.03	0.01	0.02	-0.06	0.01	-0.03	0.03
Cynosurus cristatus	0.64	0.02	-0.05	0.02	-0.04	0.02	-0.05	0.02	-0.07	0.02	-0.05	0.02	-0.07	0.02	-0.07	0.03
Dactylis glomerata sstr	0.59	0.05	0.01	0.03	0.04	0.02	0.01	0.03	0.00	0.02	0.03	0.02	-0.02	0.03	0.06	0.04
Daucus carota	0.77	0.03	-0.14	0.08	-0.08	0.04	-0.08	0.02	-0.07	0.07	-0.10	0.04	-0.08	0.06	-0.12	0.06
Deschampsia cespitosa	0.28	0.04	0.07	0.04	0.11	0.06	0.04	0.04	0.03	0.03	0.07	0.04	0.02	0.03	0.03	0.02
Euphorbia cyparissias	0.53	0.03	0.05	0.03	0.03	0.03	0.06	0.04	0.05	0.04	0.03	0.04	0.07	0.02	0.08	0.02
Euphrasia minima	0.67	0.02	0.02	0.02	-0.05	0.03	0.04	0.02	-0.06	0.03	0.01	0.02	0.02	0.02	0.01	0.03
Festuca pratensis sstr	0.49	0.02	0.03	0.03	0.02	0.02	0.03	0.02	0.04	0.03	0.00	0.03	0.00	0.02	0.01	0.03
Festuca rubra aggr	0.42	0.04	0.02	0.06	-0.01	0.04	0.04	0.03	-0.02	0.01	-0.01	0.05	-0.05	0.07	0.02	0.03
Festuca violacea aggr	0.82	0.02	0.02	0.03	0.02	0.03	0.00	0.03	0.02	0.02	0.00	0.04	0.03	0.01	0.01	0.02
Galium album	0.43	0.03	0.00	0.04	0.05	0.04	0.09	0.05	0.01	0.03	0.02	0.02	-0.01	0.02	0.07	0.04
Galium anisophyllon	0.65	0.03	0.02	0.03	-0.03	0.05	0.00	0.02	0.00	0.03	-0.01	0.02	0.01	0.03	0.01	0.01
Galium pumilum	0.32	0.02	0.01	0.04	0.05	0.04	-0.02	0.06	0.00	0.04	-0.01	0.03	0.00	0.05	0.02	0.06
Gentiana campestris sstr	0.61	0.03	0.07	0.01	0.03	0.03	0.00	0.02	0.01	0.01	0.02	0.02	0.04	0.03	-0.03	0.04
Gentiana lutea	0.48	0.06	0.01	0.01	0.01	0.03	0.06	0.03	0.01	0.02	0.08	0.04	-0.04	0.04	0.10	0.04
Gentiana purpurea	0.59	0.03	-0.03	0.01	0.02	0.03	0.06	0.03	-0.05	0.04	-0.04	0.01	-0.02	0.02	0.03	0.02
Geranium sylvaticum	0.39	0.00	-0.03	0.02	-0.07	0.01	0.04	0.03	0.00	0.01	0.00	0.02	-0.01	0.01	0.01	0.01
Geum montanum	0.64	0.03	0.05	0.07	0.05	0.03	0.17	0.06	0.01	0.02	0.04	0.03	-0.01	0.01	0.03	0.06
Helianthemum nummularium sl	0.43	0.04	0.04	0.03	0.10	0.04	0.05	0.03	0.04	0.03	0.00	0.02	0.03	0.02	0.10	0.02
Heracleum sphondylium sl	0.43	0.05	-0.04	0.03	-0.03	0.02	0.00	0.03	-0.05	0.02	0.00	0.04	-0.07	0.01	0.02	0.02
Hieracium lactucella	0.43	0.06	0.00	0.03	0.01	0.02	0.15	0.03	-0.05	0.02	0.01	0.02	0.00	0.02	-0.01	0.00
Hieracium murorum aggr	0.34	0.03	0.04	0.03	0.01	0.01	0.03	0.06	0.08	0.05	0.01	0.03	0.08	0.02	0.01	0.04
Hippocrepis comosa	0.50	0.06	-0.04	0.01	0.00	0.05	-0.01	0.03	-0.12	0.06	0.08	0.06	0.10	0.07	0.03	0.02
Holcus lanatus	0.73	0.05	0.05	0.03	0.01	0.05	0.01	0.04	0.03	0.03	0.01	0.02	0.03	0.02	-0.04	0.03
Homogyne alpina	0.57	0.03	0.05	0.02	0.02	0.03	0.00	0.02	0.00	0.03	0.01	0.03	-0.03	0.01	-0.02	0.02
Hypericum maculatum sstr	0.32	0.03	0.04	0.05	0.01	0.02	0.07	0.07	-0.03	0.03	-0.02	0.04	0.00	0.04	-0.02	0.02
Knautia dipsacifolia sstr	0.49	0.05	0.04	0.04	0.13	0.03	0.08	0.09	0.00	0.03	0.14	0.06	0.11	0.04	0.14	0.03
Lathyrus pratensis	0.58	0.02	-0.01	0.02	0.02	0.01	0.00	0.02	0.04	0.01	0.00	0.02	-0.04	0.02	0.04	0.02
Leontodon autumnalis	0.39	0.03	0.00	0.04	-0.07	0.06	0.00	0.04	-0.06	0.03	-0.03	0.07	-0.08	0.05	0.01	0.03
Leontodon helveticus	0.59	0.05	0.04	0.04	0.00	0.03	0.14	0.02	-0.02	0.04	0.00	0.03	-0.03	0.01	0.03	0.02
Leontodon hispidus aggr	0.27	0.03	-0.03	0.02	0.01	0.03	0.04	0.03	-0.05	0.01	-0.01	0.02	0.01	0.01	-0.04	0.01
Leucanthemum vulgare aggr	0.43	0.02	-0.06	0.02	-0.01	0.02	-0.02	0.01	-0.03	0.03	-0.07	0.02	-0.04	0.02	-0.09	0.01
Ligusticum mutellina	0.62	0.02	0.05	0.03	0.02	0.01	0.03	0.02	0.06	0.02	-0.01	0.02	-0.01	0.02	0.04	0.01
Linum catharticum	0.44	0.03	0.01	0.04	0.02	0.06	0.11	0.07	0.00	0.06	-0.01	0.04	-0.02	0.06	0.05	0.01

Lolium perenne	0.66	0.03	-0.01	0.05	-0.02	0.07	-0.03	0.05	0.02	0.04	0.01	0.04	-0.03	0.04	-0.01	0.05
Lotus corniculatus sl	0.28	0.02	0.04	0.02	0.02	0.03	0.04	0.01	0.04	0.02	0.02	0.04	0.04	0.01	0.05	0.03
Luzula multiflora	0.39	0.02	0.00	0.06	-0.04	0.04	0.08	0.03	0.01	0.04	0.00	0.03	-0.03	0.03	-0.01	0.02
Nardus stricta	0.57	0.03	-0.01	0.02	0.04	0.03	0.22	0.02	0.04	0.04	0.03	0.03	0.05	0.04	-0.02	0.02
Phleum pratense	0.56	0.03	-0.03	0.01	0.06	0.04	0.02	0.03	0.01	0.02	0.03	0.02	0.01	0.02	0.01	0.02
Phleum rhaeticum	0.61	0.02	0.00	0.02	-0.01	0.04	-0.03	0.02	-0.01	0.02	-0.02	0.03	-0.03	0.03	-0.01	0.02
Phyteuma orbiculare	0.37	0.04	-0.01	0.01	0.04	0.01	0.07	0.04	-0.03	0.05	-0.02	0.03	0.02	0.02	-0.09	0.03
Phyteuma spicatum	0.44	0.05	-0.04	0.02	-0.01	0.03	0.01	0.07	-0.03	0.06	-0.03	0.06	-0.03	0.03	-0.05	0.03
Pimpinella major	0.49	0.04	-0.01	0.01	0.01	0.04	-0.08	0.02	0.03	0.02	-0.01	0.01	-0.04	0.03	-0.02	0.03
Plantago alpina	0.66	0.01	0.00	0.01	0.01	0.02	0.01	0.01	-0.05	0.03	0.01	0.01	-0.01	0.01	0.00	0.02
Plantago atrata sstr	0.64	0.02	0.00	0.02	-0.02	0.03	-0.01	0.03	0.03	0.02	0.01	0.02	-0.01	0.02	-0.05	0.03
Plantago lanceolata	0.61	0.03	0.02	0.01	0.01	0.02	0.01	0.02	0.02	0.03	0.01	0.02	0.00	0.02	-0.03	0.01
Plantago media	0.51	0.06	-0.02	0.02	-0.03	0.01	0.00	0.02	-0.06	0.02	-0.06	0.01	0.00	0.05	-0.02	0.02
Poa alpina	0.66	0.01	0.02	0.02	-0.04	0.02	-0.02	0.02	-0.02	0.02	0.01	0.02	-0.05	0.02	-0.03	0.01
Polygala vulgaris	0.35	0.05	-0.01	0.03	0.01	0.03	0.03	0.03	0.14	0.01	-0.01	0.02	0.04	0.04	0.05	0.01
Polygonum bistorta	0.45	0.04	0.10	0.02	-0.02	0.03	0.05	0.04	0.08	0.03	0.00	0.04	-0.01	0.01	-0.01	0.04
Polygonum viviparum	0.58	0.02	0.01	0.01	-0.01	0.01	0.00	0.02	-0.01	0.03	-0.05	0.01	-0.03	0.01	-0.05	0.01
Potentilla aurea	0.51	0.01	0.02	0.04	0.04	0.04	0.13	0.02	-0.01	0.02	0.03	0.03	0.04	0.00	0.03	0.04
Potentilla erecta	0.50	0.02	0.00	0.03	-0.04	0.02	0.04	0.02	-0.02	0.01	0.00	0.02	0.00	0.04	-0.04	0.03
Primula elatior sstr	0.24	0.06	0.01	0.04	0.02	0.07	0.02	0.08	0.00	0.07	-0.01	0.08	-0.02	0.03	-0.09	0.04
Prunella grandiflora	0.58	0.03	-0.01	0.02	0.13	0.04	0.04	0.02	0.06	0.01	0.05	0.05	0.03	0.04	-0.01	0.01
Prunella vulgaris	0.48	0.04	0.02	0.00	0.07	0.02	0.06	0.01	0.02	0.01	-0.02	0.01	-0.01	0.03	0.00	0.01
Pulsatilla alpina sstr	0.57	0.02	-0.02	0.03	0.02	0.03	0.02	0.04	0.01	0.05	0.03	0.02	0.10	0.03	0.01	0.04
Ranunculus acris sl	0.47	0.02	0.08	0.03	0.04	0.01	-0.02	0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.00	0.01
Ranunculus montanus	0.66	0.02	-0.02	0.01	-0.07	0.01	-0.07	0.02	-0.04	0.02	-0.07	0.02	-0.03	0.03	-0.03	0.02
Ranunculus nemorosus	0.27	0.03	0.00	0.02	0.09	0.03	0.08	0.03	0.03	0.01	0.03	0.03	0.03	0.02	0.03	0.02
Rhinanthus alectorolophus	0.25	0.04	0.05	0.06	0.02	0.03	0.01	0.05	0.02	0.03	0.09	0.03	0.02	0.07	0.02	0.07
Rumex acetosa	0.57	0.01	0.04	0.02	0.02	0.01	0.04	0.01	0.01	0.01	0.02	0.00	0.03	0.02	-0.01	0.02
Sanguisorba minor	0.64	0.02	0.02	0.02	0.06	0.02	-0.01	0.01	0.01	0.05	0.02	0.01	-0.02	0.02	0.00	0.02
Scabiosa lucida	0.51	0.02	0.02	0.03	0.05	0.04	0.09	0.02	0.10	0.03	0.07	0.01	0.03	0.02	0.06	0.02
Sesleria caerulea	0.70	0.02	-0.01	0.06	0.04	0.02	0.03	0.01	-0.03	0.02	0.01	0.03	0.02	0.03	0.01	0.03
Silene vulgaris sl	0.30	0.04	0.03	0.06	0.02	0.03	0.05	0.03	0.03	0.08	0.01	0.04	-0.03	0.04	-0.01	0.09
Soldanella alpina	0.67	0.02	-0.07	0.03	-0.03	0.02	-0.08	0.03	-0.05	0.01	-0.05	0.03	-0.02	0.01	-0.04	0.03
Stellaria graminea	0.54	0.02	0.04	0.03	-0.03	0.02	0.05	0.07	-0.07	0.02	-0.01	0.04	-0.06	0.03	-0.03	0.02
Taraxacum officinale	0.53	0.03	-0.03	0.03	-0.01	0.02	-0.02	0.02	-0.03	0.01	-0.02	0.03	0.00	0.02	-0.03	0.02
Thymus praecox ssp polytrichus	0.45	0.04	-0.12	0.07	-0.06	0.06	-0.07	0.04	-0.09	0.02	-0.12	0.03	-0.07	0.02	-0.05	0.03
Thymus pulegioides sstr	0.51	0.04	0.01	0.02	-0.05	0.02	-0.06	0.01	-0.04	0.02	0.00	0.02	-0.07	0.03	-0.07	0.02

Tragopogon pratensis	0.44	0.06	0.00	0.03	-0.05	0.03	-0.03	0.06	-0.09	0.05	-0.03	0.02	-0.06	0.01	-0.07	0.04
Trifolium badium	0.49	0.03	-0.04	0.03	-0.02	0.02	-0.04	0.03	-0.04	0.05	-0.04	0.03	0.01	0.04	0.02	0.02
Trifolium medium	0.51	0.02	0.01	0.03	0.01	0.06	-0.04	0.04	0.07	0.05	-0.06	0.02	0.00	0.04	0.04	0.03
Trifolium pratense sstr	0.36	0.03	0.00	0.04	-0.03	0.02	0.04	0.03	0.00	0.04	-0.03	0.01	0.03	0.03	-0.03	0.05
Trifolium repens sstr	0.51	0.02	0.03	0.02	0.04	0.02	0.07	0.01	0.04	0.01	0.04	0.01	-0.01	0.03	0.01	0.02
Trisetum flavescens	0.47	0.03	0.07	0.02	0.02	0.02	0.02	0.04	0.02	0.03	0.02	0.03	0.01	0.03	0.04	0.03
Trollius europaeus	0.37	0.02	-0.02	0.03	0.08	0.02	-0.02	0.03	0.00	0.01	0.05	0.04	0.06	0.03	0.09	0.03
Vaccinium myrtillus	0.43	0.03	0.02	0.03	-0.06	0.05	0.27	0.02	0.02	0.05	0.06	0.06	0.03	0.05	-0.01	0.05
Veratrum album ssp lobelianum	0.59	0.05	0.04	0.05	0.03	0.02	0.02	0.01	0.03	0.01	0.06	0.04	0.01	0.01	0.01	0.03
Veronica chamaedrys	0.59	0.02	0.01	0.01	-0.02	0.03	-0.06	0.01	-0.04	0.01	-0.01	0.03	0.03	0.02	-0.03	0.01
Veronica serpyllifolia sl	0.28	0.05	0.19	0.06	0.13	0.06	-0.03	0.02	0.04	0.05	0.06	0.08	0.00	0.07	0.06	0.07
Vicia sepium	0.55	0.02	-0.01	0.01	-0.01	0.01	0.01	0.03	-0.02	0.01	0.03	0.03	-0.04	0.03	-0.02	0.03

Table S3. Pearson correlation values between all topo-climatic and edaphic variables used as predictors in the models.

	N	pH	P	Cl	Sa	Si	Degree day	Slope	Moisture	Radiation	Topo
C:N	-0.04	0.25	-0.09	-0.25	-0.13	-0.34	-0.11	0.05	0.15	-0.10	-0.02
N		0.05	0.43	0.41	-0.28	0.36	-0.17	-0.18	0.05	0.09	0.18
pH			0.07	-0.11	-0.13	-0.25	0.00	0.15	0.02	0.06	-0.09
P				0.25	-0.18	0.21	-0.09	-0.23	-0.03	0.04	0.10
Cl					-0.70	0.91	0.09	-0.10	-0.20	0.23	0.02
Sa						-0.70	-0.04	0.06	0.10	-0.08	0.04
Si							0.13	-0.10	-0.18	0.12	0.01
Degree day								-0.26	-0.83	0.06	-0.57
Slope									0.33	-0.10	0.22
Moisture										-0.43	0.45
Radiation											0.09

Table S4. Pearson correlation coefficients quantifying the relation between the changes in AUC caused by the addition of each edaphic variable for each pair of modelling technique.

	GAM-GBM	GAM-GLM	GAM-RF	GBM-GLM	GBM-RF	GLM-RF
pH	0.753	0.855	0.698	0.787	0.896	0.739
N	0.670	0.754	0.754	0.604	0.849	0.596
C:N	0.646	0.786	0.627	0.659	0.872	0.589
P	0.625	0.697	0.558	0.644	0.833	0.519
Si	0.658	0.760	0.646	0.569	0.807	0.522
Cl	0.546	0.722	0.492	0.600	0.820	0.502
Sa	0.495	0.727	0.389	0.616	0.816	0.504

Figure S1

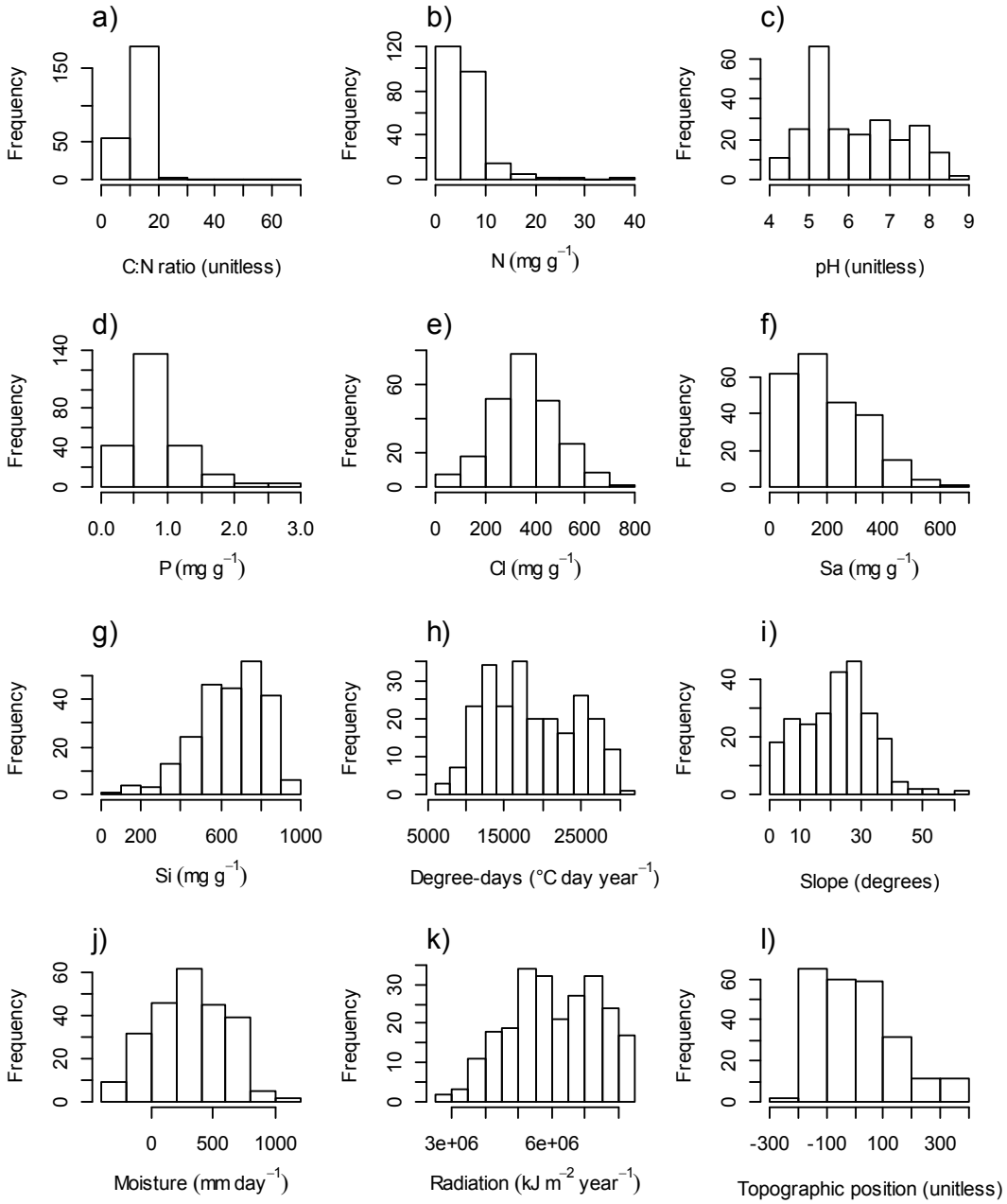


Fig. S1. Histograms of the distribution of the predictors used to calibrate the models. a) carbon to nitrogen ratio, b) nitrogen content, c) pH, d) phosphorus content, e) clay content, f) sand content, g) silt content, h) degree days, i) slope, j) moisture index, k) solar radiation, l) topographic position.

Figure S2

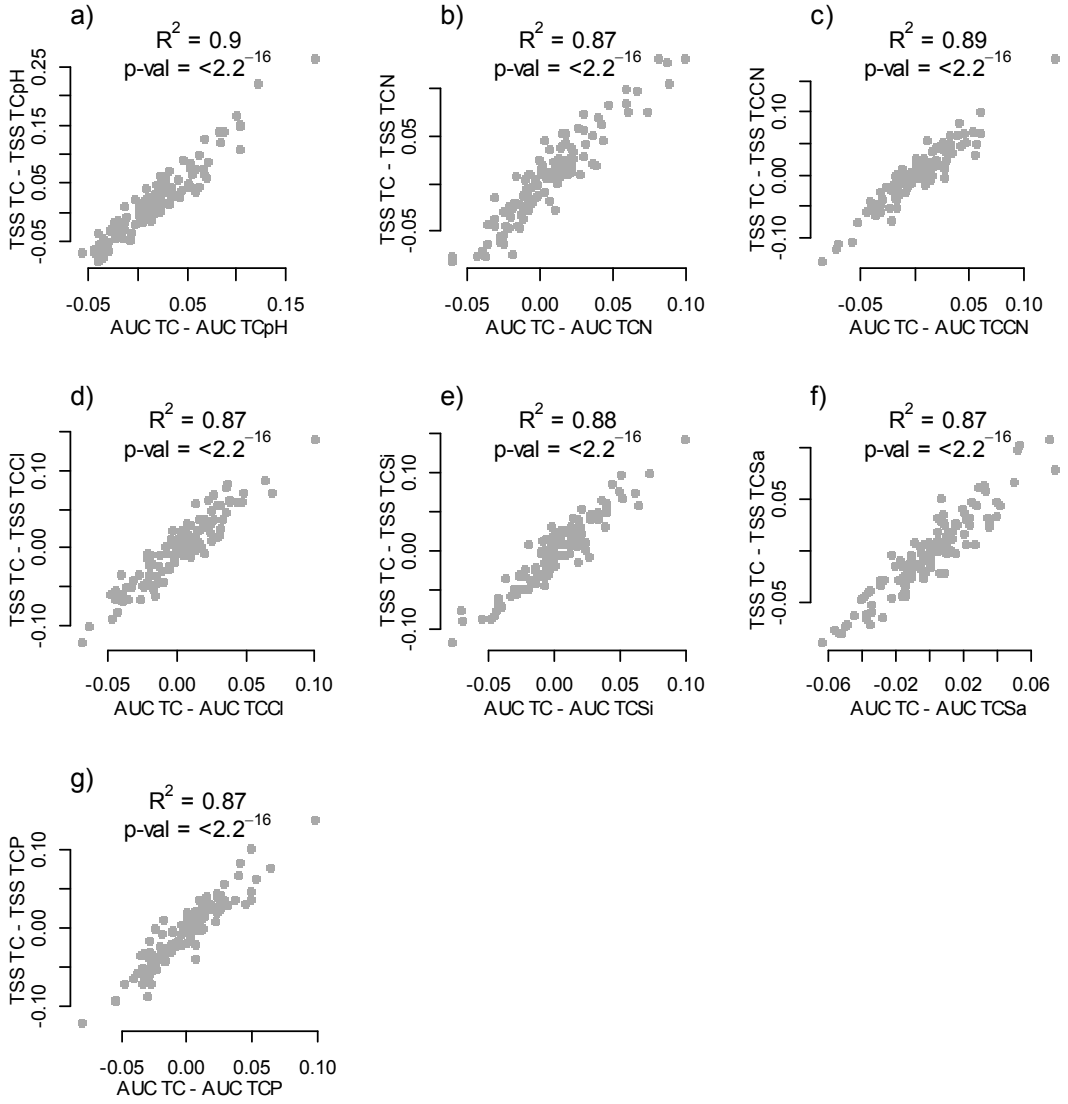


Fig. S2. The relationship between the change in the AUC and TSS values (average of the four modelling technique) caused by adding each edaphic variable in the distribution models for each species, a) for pH, b) for nitrogen content, c) for carbon to nitrogen ratio, d) for clay content, e) for silt content, f) for sand content, g) for phosphorus content. The correlation coefficients range from 0.93 to 0.95.

Figure S3

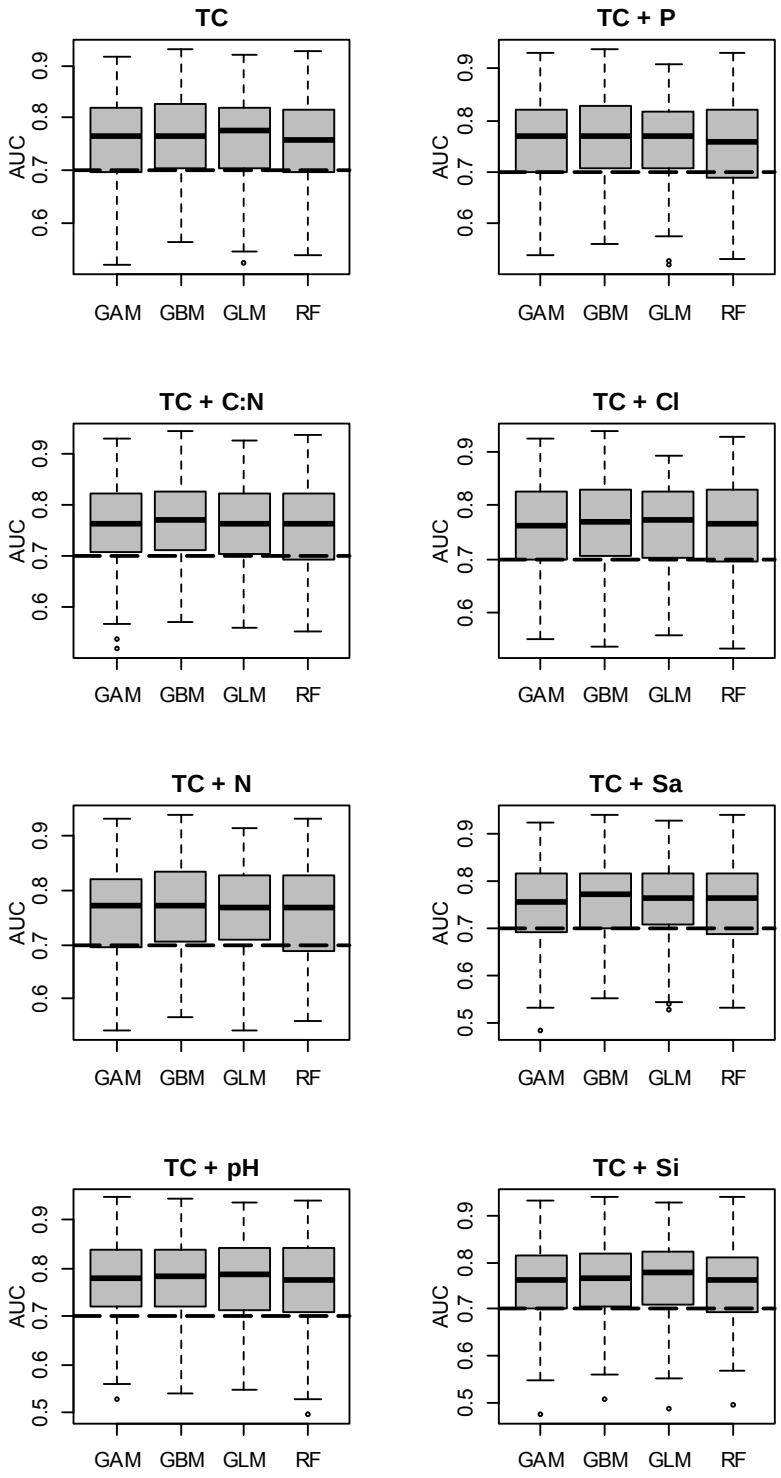


Fig. S3. Boxplot of the AUC for all species sorted by modelling techniques, for each combination of topo-climatic and environmental predictors. Models, for which the AUC is higher than 0.7 (horizontal line), are considered as reliable.

Figure S4

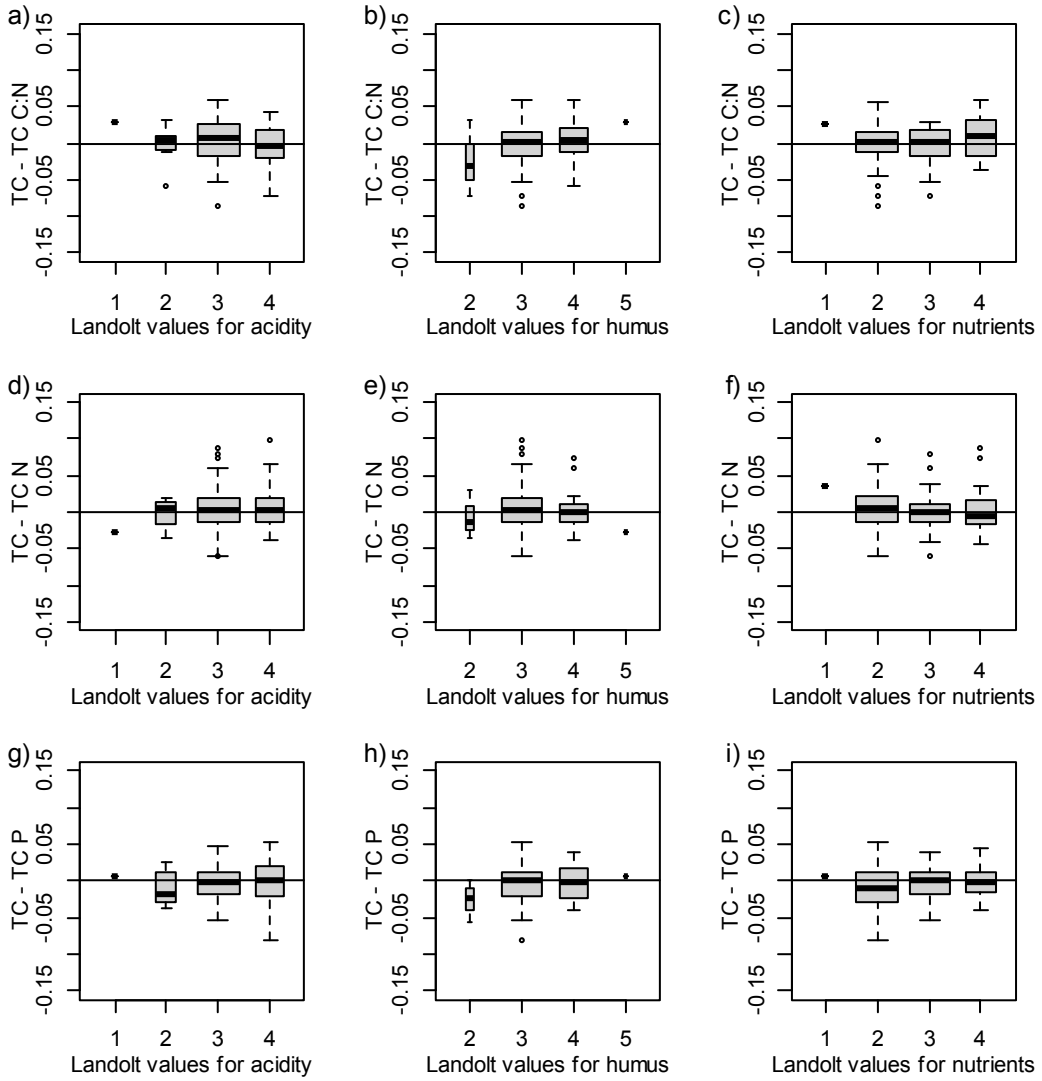

a)
b)
c)
d)
e)
f)
g)
h)
i)
TC - TC C:N
TC - TC N
TC - TC P
-0.15
-0.05
0.05
0.15
1
2
3
4
5
Landolt values for acidity
Landolt values for humus
Landolt values for nutrients

Figure S4 (continued)

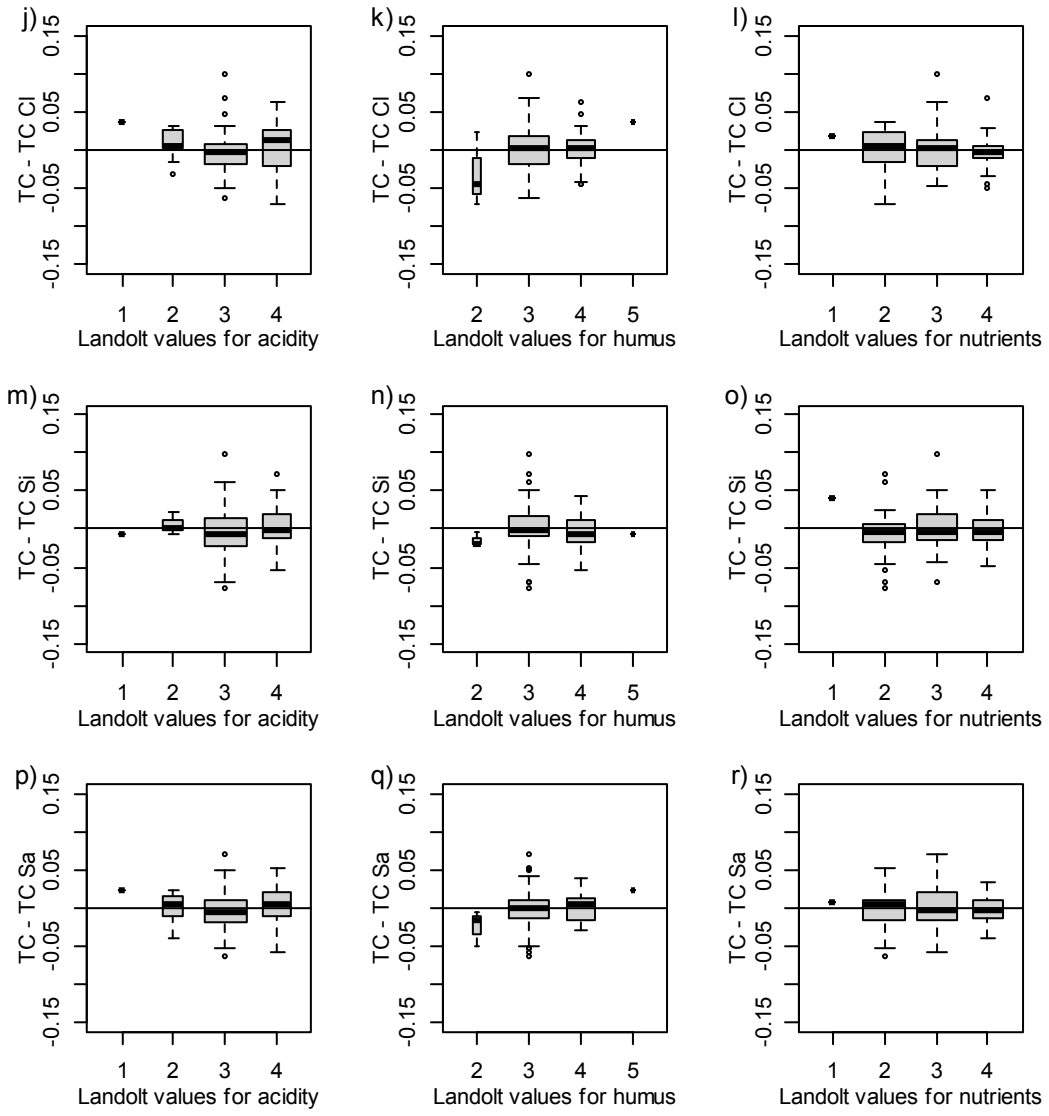


Fig. S4. The change in the AUC upon the addition of edaphic predictors a),b),c) carbon to nitrogen ratio, d), e), f) nitrogen content, g), h), i) phosphorus content, j), k), l) clay content, m), n), o) silt content, p), q), r) sand content; compared to topo-climatic models, for each species (mean of the results for the four modelling techniques). The plants are classified according to their preferred soil conditions as in Landolt et al. (2010). The width of the box-plots is proportional to the number of species taht show the indicator value of interest. The middle line is the median, and the outer limits of the box-plots are the 1st and 4th quartiles. For these cases no clear trends can be detected.

Figure S5

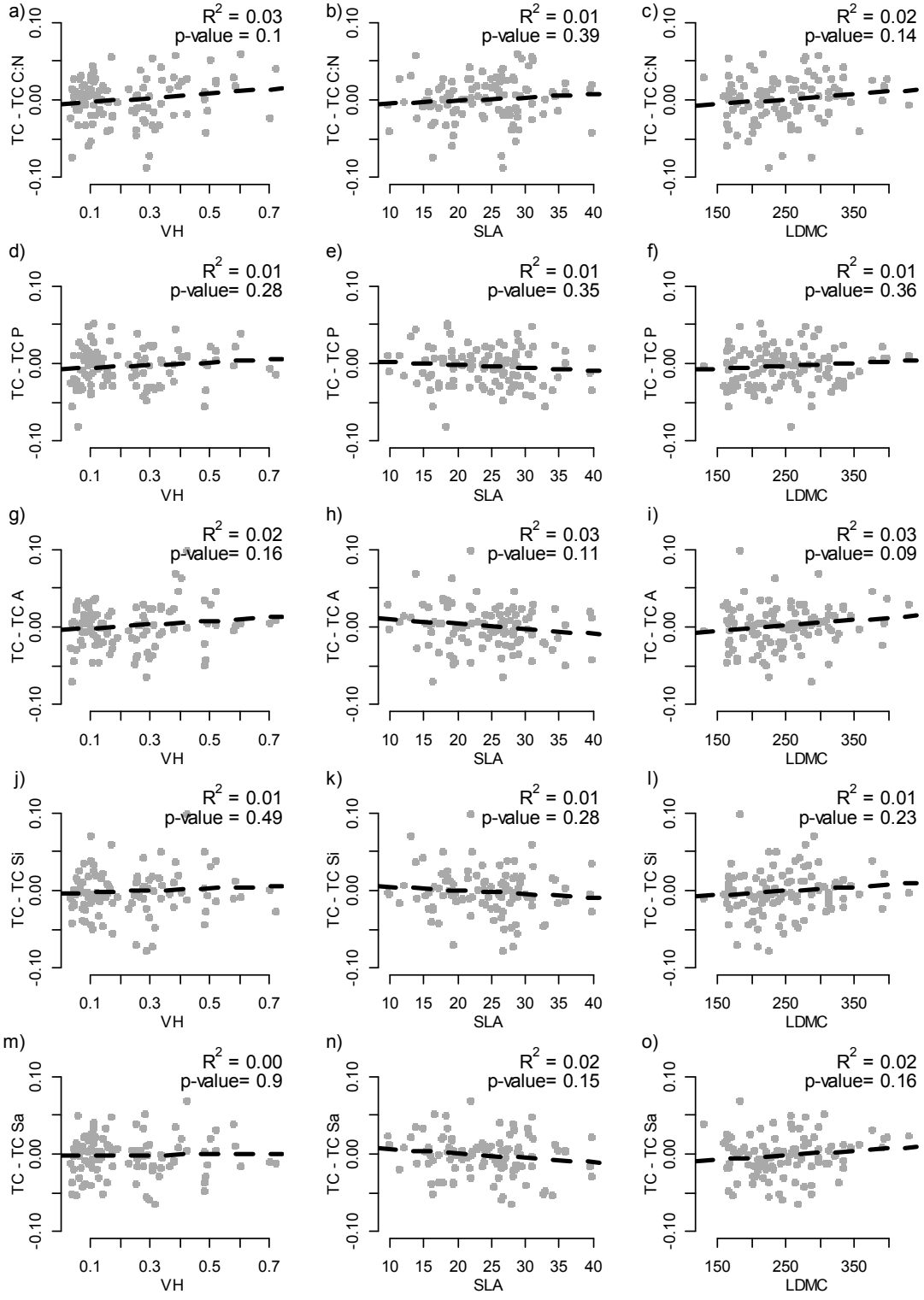


Fig. S5. The mean AUC change between the topo-climatic models, with or without the edaphic predictors: a),b),c) carbon to nitrogen ratio, d), e), f) phosphorus content, g), h), i) clay content, j), k), l) silt content, m), n), o) sand content; for each species plotted against their height (VH, in meter, first column), specific leaf area (SLA, in mm^2*g^{-1}, second column) and leaf dry matter content (LDMC, in $mg*g^{-1}$, third column). The dashed line is the linear regression. In these cases, no significant relationship could be evidenced.

CHAPTER 2

Very high-resolution environmental predictors in species distribution models: moving beyond topography

Jean-Nicolas Pradervand [1]

Anne Dubuis [1]

Loïc Pellissier [1]

Christophe Randin [2]

Antoine Guisan [1,3]

1 Department of Ecology and Evolution, University of Lausanne, Biophore building, 1015 Lausanne, Switzerland

2 Botanisches Institut der Universität Basel, Schönbeinstrasse 6, 4056 Basel Switzerland

3 Faculty of Geoscience and the Environment, University of Lausanne, 1015 Lausanne, Switzerland

This paper is in preparation for Progress in Physical Geography

Contribution to the project: I participated to the field work. I performed the statistical analysis with J-NP and we wrote the paper together. Both first authors contributed equally

ABSTRACT

Recent advances in remote sensing technologies have facilitated the generation of very high-resolution (VHR) environmental data. If used in species distribution models (SDMs), these data could enable modelling the species micro-habitats and allow improved predictions to be made for fine-scale biodiversity management.

In the present study, we tested the influence of the predictors resolution in SDMs by comparing the predictive power of models for 239 plant species and their assemblages fitted at six different resolutions (from 2 m to 100 m) in a 700 km^2 study area in the Swiss Alps. We also tested whether changes of the model quality for a species is related to its functional and ecological characteristics.

Refining the resolution contributed to slightly improving the models for more than half of the examined species, with the best results obtained at 5 m. For the other species, no changes could be detected, or the best model remained the one with the coarsest resolution, again with little change in predictive power being observed. Contrary to our expectations, we could not consistently correlate the changes in model performance with species characteristics.

Temperature was the most important variable in the SDMs across the different resolutions. However, maps of modelled temperatures at higher resolutions (2-10 m) than previously used (25-100 m) did not picture enough variation and could not substantially improve the predictive power of the SDMs. This is likely because variables other than temperature (e.g., soil moisture, snow cover) take on more importance at a micro-scale, but accurate measurements are still lacking for these variables.

Synthesis and applications. Our results suggest that improving resolution of topographic data only, without also conducting finer-scale in-situ environmental measurements (e.g. for temperature, moisture, snow), is not sufficient to improve species prediction models - and therefore local management - compared to previously used resolution (here 25 and 100 m). One way to obtain improved environmental maps at a finer scale would be to use field loggers or remote sensing technology. More effort should be dedicated now to obtaining improved environmental measurements for fine-scale species mapping and management..

Key-words: Species distribution models (SDMs); plants; grain size; very high resolution; microtopography, mountains

INTRODUCTION

Species distribution models (SDMs) are widely used tools for understanding and predicting spatial patterns of biodiversity. By relating species occurrences with environmental variables, these models allow quantification of species niches, testing of ecological hypotheses and prediction of species distributions under past, present and future environmental conditions (Guisan & Zimmermann 2000; Guisan & Thuiller 2005; Elith & Leathwick 2009). SDMs are thus a powerful tool for species and biodiversity management (Elith & Leathwick 2009; Falk & Mellert 2011). The quality and ecological significance of the environmental variables used as predictors in the models are very important (Austin 2007; Austin & Van Niel 2011a), especially if communities and ecosystems have to be predicted by stacking individual species (Guisan & Rahbek 2011), as even small errors in individual models can accumulate into larger errors in assemblage predictions (Dubuis *et al.* 2011; Pottier *et al.* 2013). Randin *et al.* (2009b), Coudun *et al.* (2006), Bertrand *et al.* (2012) and Dubuis *et al.* (2012) demonstrated the need to include land use and soil descriptors, while Randin *et al.* (2009c) and Le Roux, Virtanen, & Luoto (2013) highlighted the importance of including geomorphological disturbances in species distribution models. Surprisingly however, in comparison to the tremendous increase in the peer-review literature on SDMs, a relatively low effort has been invested toward improving the predictors used in models. Without first progressing significantly in our capacity to map the fine-scale abiotic requirements of species (Austin & Van Niel 2011b), we may not be able to assess what limits our ability to predict species distributions.

An intuitive concept when modelling low-stature plants in harsh environments is that fine-scale descriptors of the environment would be required (Lassueur, Joost, & Randin 2006). Climate acts as a filter, allowing the growth of species displaying suitable physiological traits and is expected to be a main determinant of the distribution of most species (Woodward 1987). Topography can act either indirectly, by modulating micro-climatic and soil properties (e.g., solar radiation, drainage), or directly, through slope and associated gravitational processes. Species occupying micro-environments, i.e., for which some dimensions of their niche are defined by, for example, the fine-scale topography, are especially in need of more detailed climatic and topographic predictors. For instance, alpine species usually occupy microclimatic environments that may be decoupled from general conditions at coarser resolutions (e.g., ground versus air temperature; Randin *et al.* 2009d; Austin & Van Niel 2011a; Scherrer & Körner 2011; Pellissier *et al.* 2013). These same micro-environments may act as refugia for cold-

adapted species under climate change (Randin *et al.* 2009a). To achieve more realistic forecasts for species distributed in microenvironments, it therefore seems particularly important to improve the quality and resolution of the environmental variables used in SDMs. Very high-resolution (VHR) predictors could prove particularly interesting in mountain areas with a rugged topography, where temperatures may vary by several degrees across a few metres and within a few hours during the day (Scherrer, Schmid, & Körner 2011). This situation highlights the importance in such environments of using the most precise and accurate predictors available to approximate the fine environmental requirements of species as closely as possible.

Recent advances in remote sensing technologies and computer performance have opened the way for the creation of VHR variables and digital elevation models (DEMs) below the scale of metres (Ehlers, Gaehler, & Janowsky 2006). These VHR DEMs have been applied in engineering studies for high-precision geomorphological analyses, such as flooding calculations (Vaze, Teng, & Spencer 2010; Kalbermatten *et al.* 2012), but they could also be employed to improve the current topographic or environmental variables used in SDMs. VHR predictors might, however, increase the precision and accuracy of SDMs through better capturing micro-environmental variation, providing a level of information that was previously inaccessible to modellers or only accessible across very limited geographic extents (Gottfried *et al.* 1999; Lassueur, Joost, & Randin 2006; Wang *et al.* 2012; Le Roux *et al.* 2013).

With the rise of VHR DEMs, it has therefore become an important challenge to test whether an increase in the resolution of a DEM and the associated variables (e.g., topographic and climatic variables interpolated at a VHR) allows improvement of the predictive power of SDMs compared to models computed at a standard resolution (e.g., 25 m, as employed in a mountain landscape by Dubuis *et al.* 2011).

The usefulness of VHR maps as predictors in SDMs has been poorly studied due to the associated cost and data availability but also to the difficulty of projecting SDMs over large areas at a fine resolution. Several studies have compared the performance of SDMs between two or more different grain sizes, but always at a coarse resolution (e.g., 100 m has been the finest resolution used) or for a reduced number of species. In these studies, the effect of the resolution on SDMs for plants across regions, species and techniques was found to be weak (Guisan *et al.* 2007b; a). Lassueur *et al.* (2006) showed that VHR topographic predictors can improve SDMs, but their study only tested VHR predictors in linear models without quadratic terms and therefore without unimodal response curves. Some pioneering works coupling the recent advances in numerical

mapping (e.g., high-resolution elevation models, hyperspectral images, thermal imagery and field mapping) with species mapping have shown great promise (Camathias *et al.* 2013). We now have the capacity to conduct similar studies over large areas, such as entire regions of the Alps, or at multiple scales, a required condition to answer certain important questions (McGill 2010; Nogués-Bravo & Rahbek 2011).

Here, we test whether using topographic and temperature predictors directly derived from a DEM at a very high resolution (2-10 m) improves the predictive power of SDMs for 239 plant species and the resulting species assemblage compared to previous models fitted with similar predictors, but at lower resolutions (25-100 m). Second, we ask if the optimal resolution for a species is related to some of its functional traits. Due to the rather low stature of most plant species considered, we expected to observe i) an improvement of SDMs and associated assemblage predictions with an increasing resolution of environmental predictors and ii) the models to exhibit different predictive powers across the tested resolutions, with the optimum depending on the traits of the species involved.

METHODS

Study area and species data

The study area was located in the western Swiss Alps (Canton de Vaud, Switzerland, 46°10' to 46°30'N; 6°50' to 7°10'E). It covered ca. 700 km^2, with an elevation gradient stretching from 375 m asl to 3210 m asl. For more information on this area, see Randin et al. (2009b). The species occurrence data used in our analysis originated from fieldwork conducted between 2002 and 2009 in the study area (Dubuis *et al.* 2011) following a random-stratified sampling design (Hirzel & Guisan 2002) based on elevation, slope and aspect. The sampling was limited to open, non-woody vegetation, and each selected point was separated from the others by a minimum distance of 200 m to avoid spatial autocorrelation. For the initial dataset, 613 vegetation plots of 4 m^2 each were inventoried, and the resulting data were used for SDM calibration. An additional set of 298 plots was surveyed identically to evaluate various SDMs. This evaluation dataset can be considered as independent of the first one (Pottier *et al.* 2013). The position of each plot was recorded using a Trimble GEO Explorer GPS allowing a submetric accuracy. The vegetative height (H) was measured in the field (in mm) as the distance between the top of the photosynthetic tissue and the ground. This trait is related to competitive ability and is correlated with the above-ground biomass. The maximum of elevational range was noted for each species.

Environmental predictor variables

As a basis for computing predictor variables, we used digital elevation models (DEM) at six different resolutions: 2, 5, 10, 25, 50, and 100 m. Among these DEMs, the 25 m and 100 m DEMs were obtained from independent sources from the Federal Office of Topography (Swisstopo). The 2, 5, and 10 m DEMs were computed by aggregating a 1 m DEM (Swisstopo), and the 50 m DEM was computed by aggregating the 25 m DEM. The minimum resolution was determined by the resolution of the vegetation plots (2x2 m). These DEMs served as the basis for computing the following data.

Topographic position: This predictor is an integration of topographic positions discriminating convex situations (ridges, bumps) from regular slopes (mountain sides, flat areas) and from concave situations (valley bottoms, depressions), calculated at various scales. It was computed using an ArcInfo Macro Language custom code in ArcGIS 10 for each DEM resolution by means of multi-radius moving windows (see Randin *et al.* 2009a; b for more details). For the 2, 5, and 10 m resolutions, topographic

features were detected for each cell of the landscapes within a window radius with increments ranging from 10 m to a 200 m radius. For the 50 and 100 m DEMs, increments of 50 and 100 m were used, respectively.

Slope: The slope was calculated at each resolution with the spatial analyst tool in ArcGIS 10.

Temperature: During 2006, 50 temperature loggers were set up in the study area, recording one temperature measurement per hour. To avoid overheating problems and to capture the equilibrium with cold atmospheric conditions, we used the daily minimum temperatures. Moreover, we chose to use the temperatures obtained in June, July and August, as these months can be considered to correspond to the plant growing season in our study area and to the snow-free period. For each logger, we used the closest meteorological station of the Swiss Meteorological Institute (MeteoSwiss) as a reference station to reconstruct 30 years of temperatures (the 1981-2010 period) by hindcasting at the site of the logger. A standard period of 30 years was reconstructed to better reflect long-term temperature conditions and the variation experienced by most perennial plants and to avoid spurious correlations if the calibration year were to be abnormal (Sup mat. Fig. S6). We predicted the temperatures at the loggers based on linear models that relate the temperature of each logger to the reference weather station during the 2006 calibration year and then used this relationship to hindcast for the standard 30-year period.

We next computed the long-term monthly mean minimum temperatures for each logger for each month. Temperature maps were then derived at each resolution using a generalised linear model relating the logger reconstructed long-term temperature means with elevation and topographic position as predictors. Polynomials were allowed up to the second order and a Gaussian distribution of errors was assumed. The fit of this model was calculated with an adjusted D^2, and its predictive power was computed with the root mean square error (RMSE) after a leave-one-out cross-validation. Finally, we averaged the results for June, July and August to obtain the minimum temperature average for the plant growing season.

Solar radiation: The amount of solar radiation received in each month of the growing season (June, July and August) in each pixel was computed at each resolution with the spatial analyst tool in ArcGIS 10.

Correlations were finally computed between each resolution for each predictor to evaluate the differences in predictor values due to the change of resolution.

Species distribution models

We modelled the distribution of 239 plant species with a minimum of 15 occurrences in the calibration dataset using the average minimum temperature for the growing season, the slope, the topographic position and the sum of solar radiation for the growing season as predictors. All models were fitted in R (2.14.1) with the BIOMOD library (Thuiller *et al.* 2009). We employed three different types of models shown to provide satisfactory predictions of species distributions in a large comparative study (Elith *et al.* 2006): generalised linear models (GLMs), generalised additive models (GAMs), and generalised boosted regression models (GBMs). The GLMs and GAMs were calibrated using a binomial distribution and a logistic link function, and polynomials were allowed up to the second order for each predictor. The GBMs was calibrated with 2 000 trees. The models were calibrated based on the 612 plots in the calibration dataset, and their predictive power was evaluated using the area under the curve (AUC, Hanley & Mcneil 1982), true skills statistics (Allouche, Tsoar, & Kadmon 2006), sensitivity, and specificity in the evaluation dataset of 298 plots. The performances of the models were compared between resolutions for all species with Wilcoxon signed rank tests. We also sorted the species according to their characteristics (vegetative height and maximum elevation range) to test whether some groups showed better predictive power at particular resolutions.

Finally, for each species and each resolution, the importance of each variable in the models was assessed in BIOMOD by randomizing each variable individually and then projecting the model on the randomized variable while keeping the other variables unchanged. The results of the prediction with the randomized variable were then correlated with those of the original models.. Finally, the importance of the variable was calculated as one minus the correlation, with higher values indicating predictors that are more important for the model (Thuiller *et al.* 2009).

Species assemblage composition prediction

To test the influence of the resolution on the prediction of plant assemblages, we generated one ensemble projection per species by averaging the predictions resulting from each modelling technique weighted by its respective AUC on the evaluation dataset (Araújo *et al.* 2007). For each model, the predicted probabilities were transformed into binary presence/absence data to maximise the sensitivity and specificity of the model (Liu *et al.* 2005). By stacking the binary predictions for each

species, we were able to predict the species composition for each plot in the evaluation dataset. This stacking of model outputs was performed at each resolution.

The quality of the predicted assemblage was evaluated using several metrics based on a confusion matrix into which all of the species included in the analysis (species pool: *SP* =239) were classified as follows: *TP:* the species both observed and predicted to be present (true positive); *FN:* the species observed to be present, but predicted as absent (false negative; omission error); *FP:* the species observed to be absent, but predicted to be present (false positive; commission error); and *TN:* the species both observed and predicted to be absent (true negative). We computed the assemblage prediction success (a) and the Sørensen index (b):

$$(a)\ Prediction\ success = \frac{TP + TN}{SP}$$

$$(b)\ Sørensen\ index\ = \frac{2TP}{2TP + FN + FP}$$

We also evaluated the assemblage predictions retaining the results as probabilities and using the following metrics that allow comparison of the probabilities with binary data: the brier score (c), in which *f* is the probability that was forecast, while *o* is the actual observation (0 if the species is absent and 1 if it is present), and *SP* is the number of species included in the analysis; and the continuous form of the Sørensen index (d), in which y_i and y_j are the plots under comparison, while *k* is the index of a species, and *SP* is the total number of species.

$$(c)\ Brier\ score = \frac{1}{SP}\sum(ft - ot)^2$$

$$(d)\ Sørensen\ index = 1 - \frac{\sum_{k=0}^{SP-1}\left|y_{i,k} - y_{j,k}\right|}{\sum_{k=0}^{SP-1}\left(y_{i,k} - y_{j,k}\right)}$$

Finally, for the binary and probability predictions, the species richness (SR) error was computed as the predicted species richness minus the observed species richness.

RESULTS

Evaluation of the temperature models

The fit of the models used to spatially predict temperature was rather high, showing D^2 values between 0.76 and 0.80 (average of 0.769) and adjusted D^2 values between 0.75 and 0.78 (average of 0. 754). The complete results are presented in Table S1.

Species distribution models

Overall, the accuracy of the SDM predictions, measured with the AUC, presented an absolute range between 0.36 (counter-prediction) and 0.98 (excellent). Across all resolutions, the bulk of the predictions were fair to good, with 95% of values falling between 0.62 and 0.93 and a mean AUC value of 0.80 being obtained (Table 1, Fig. 1a).

Table 1. Means, variances and standard deviations for the AUC values (mean of the three modelling techniques for each species).

	AUC		
	Mean	**Variance**	**SD**
2 m	0.797	0.006	0.078
5 m	0.799	0.006	0.075
10 m	0.797	0.006	0.077
25 m	0.794	0.006	0.080
50 m	0.794	0.006	0.078
100 m	0.787	0.006	0.081

On average, the 5 m SDMs showed significantly higher AUC values than the 2 m and 100 m SDMs, and the 10 m SDMs also exhibited significantly higher AUC values than those at 100 m (Table 2). However, the increase in the AUC was small, with the largest improvement, of 1.12%, being observed for the 5 m models compared to the 100 m models. The sensitivity and specificity showed no marked differences between resolutions (Fig. S1, S2). The trends for the TSS were similar and are presented in Fig. S3, S4 and S5.

Table 2. Wilcoxon signed rank tests for the AUC (mean of the three modelling techniques for each species) at each resolution with the Bonferroni correction (p<=0.00083).

	Direction of the test
2 m-5 m	*Lower*
2 m-10 m	*n.s.*
2 m-25 m	*n.s.*
2 m-50 m	*n.s.*
2 m-100 m	*Higher*
5 m-10 m	*n.s.*
5 m-25 m	*n.s.*
5 m-50 m	*n.s.*
5 m-100 m	*Higher*
10 m-25 m	*n.s.*
10 m-50 m	*n.s.*
10 m-100 m	*Higher*
25 m-50 m	*n.s.*
25 m-100 m	*n.s.*
50 m-100 m	*n.s.*

Hereafter, we present the mean AUC values among the three modelling techniques. The results for each technique can be found in the supplementary information (Table S2).

Fig. 1b compares the mean AUC values obtained for the 5 m models (for the GLM, GAM and GBM) for each species with the corresponding AUC values for the 10 m, 25 m, 50 m, and 100 m models. The delta AUC varied with the differences in resolution. The 5 m-100 m difference in resolution showed the greatest mean variance. However, the median did not change across the different resolutions, remaining equal to zero, and the variations for 75% of the species remained lower than 5%, even for the 2 m – 100 m comparison (upper quartile = -0.012, lower quartile = 0.033).

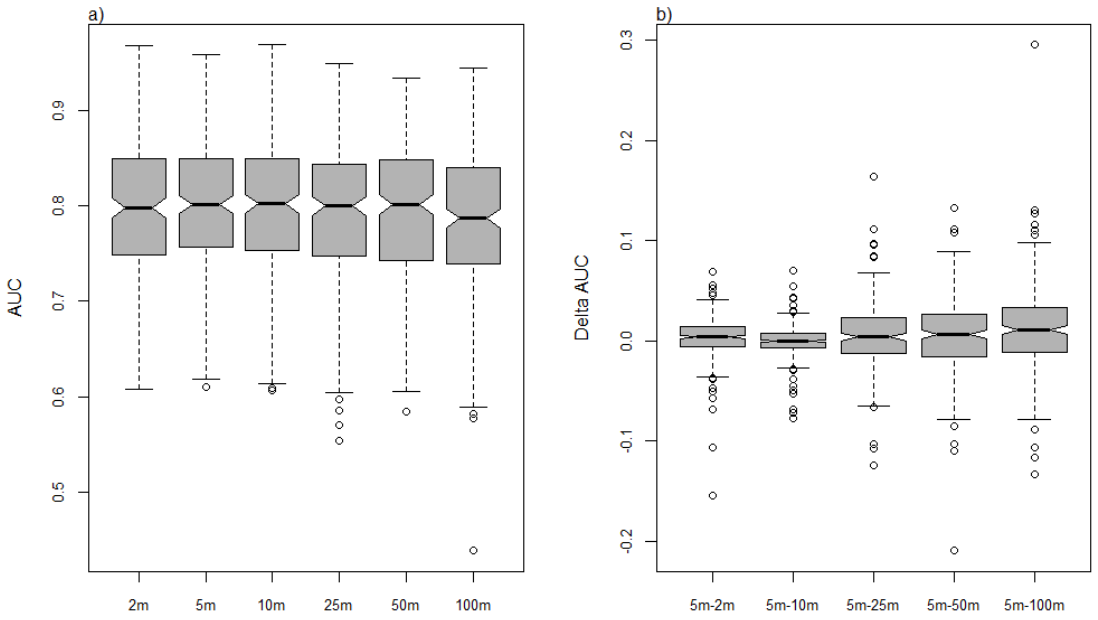


Figure 1. (a) Boxplot of the AUC values (mean of the three modelling techniques for each species) according to the different resolutions examined. (b) Comparison of the AUC values for the 5 m models (with the AUC values for the 10, 25, 50 and 100 m resolutions).

Species characteristics

The differences in the AUC values for the models did not change according to species characteristics. Fig. 2 & 3 show the delta AUC (average for the three modelling techniques) as a function of the maximal elevation and vegetation height, respectively.

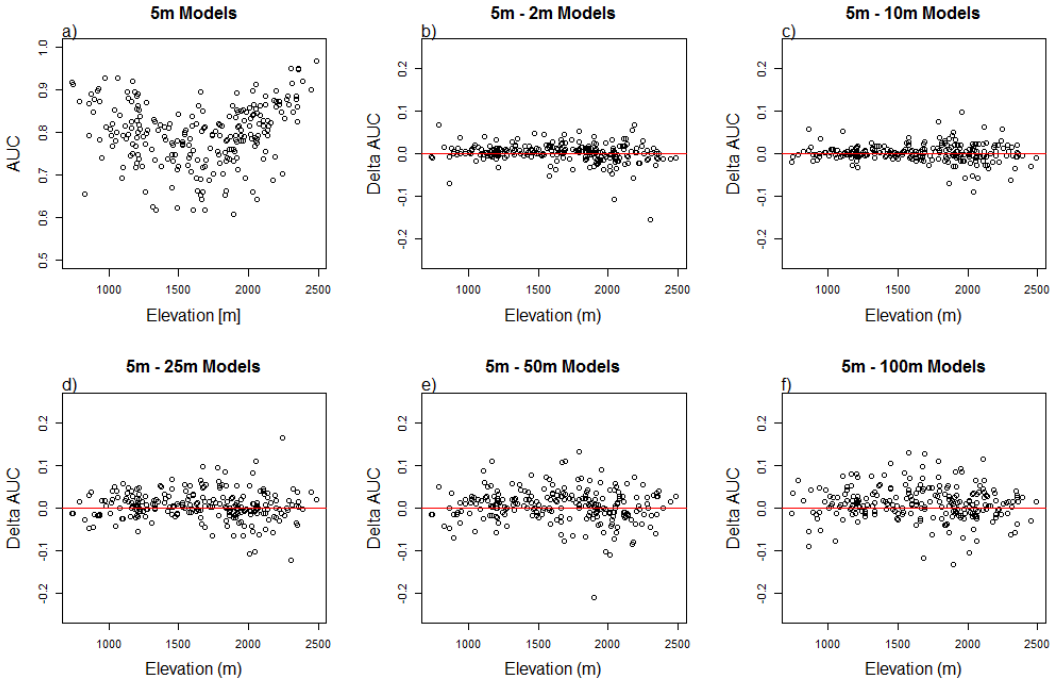


Figure 2. (a) The AUC for the 5 m models compared to the maximal elevation for each species, (b-f) delta AUC between the 5 m AUC values and the AUC values for the coarser resolutions for each species plotted along the maximal elevation for the species. The line corresponds to a delta AUC of zero.

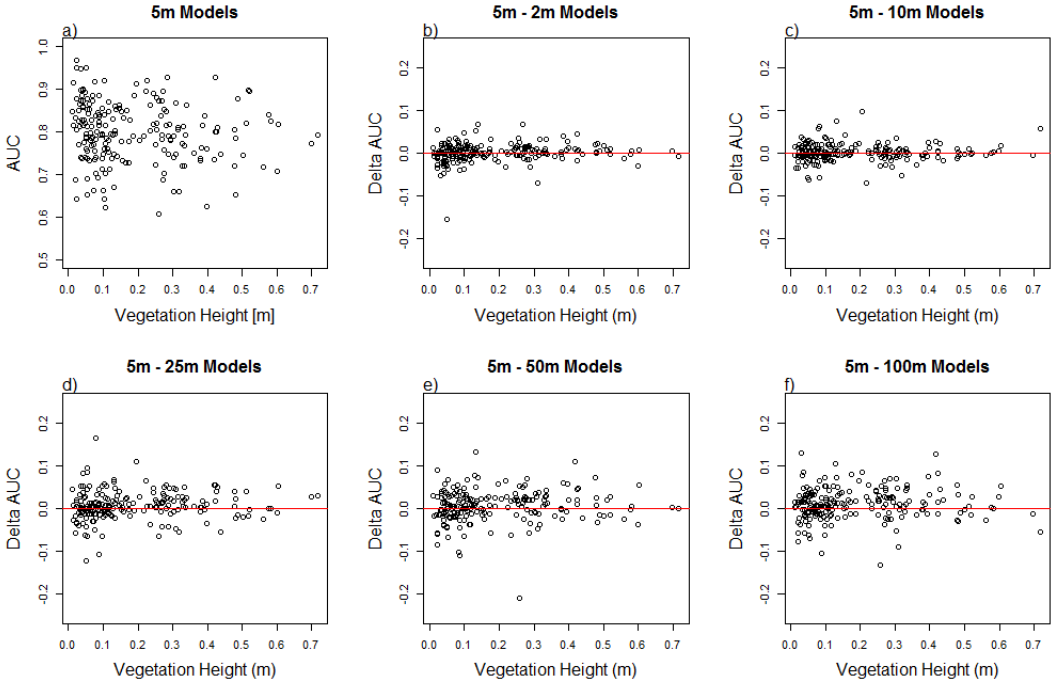


Figure 3. (a) The AUC of the 5 m models as a function of the maximal height for each species, (b-f) delta AUC between the 5 m AUC values and the AUC values of the coarser resolutions for each species plotted along the vegetative heights of the species. The line corresponds to a delta AUC of zero.

Species assemblage predictions

The species richness error of the binary prediction showed a slight trend toward increasing at coarser resolutions (Table 3, Fig. 4), while the prediction success values and inverse of the Brier scores decreased significantly towards coarser resolutions. The Sorensen indices for the binary and probability predictions showed no significant differences.

Table 3. The results of the Wilcoxon tests for the communities. The first three columns show the results of the evaluation of the species assemblage predictions after transforming the ensemble forecasting into binary presences and absences: (i) species richness error, (ii) Sorensen similarity and (iii) prediction success. The next three columns display the evaluation results based on comparison of the probabilities predicted by ensemble forecasting with the observed binary data: (iv) species richness error, (v) Sorensen similarity and (vi) inverse of the Brier score.

	SR error bin	Sorensen	Prediction success	SR error proba	Sorensen proba	1-Brier score
2 m - 5 m	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
2 m - 10 m	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
2 m - 25 m	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
2 m - 50 m	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
2 m - 100 m	Lower	n.s.	Higher	n.s.	n.s.	n.s.
5 m - 10 m	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
5 m - 25 m	n.s.	n.s.	n.s.	n.s.	n.s.	Higher
5 m - 50 m	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
5 m - 100 m	Lower	n.s.	Higher	n.s.	n.s.	Higher
10 m - 25 m	n.s.	n.s.	n.s.	n.s.	n.s.	Higher
10 m - 50 m	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
10 m - 100 m	Lower	n.s.	n.s.	n.s.	n.s.	Higher
25 m - 50 m	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
25 m - 100 m	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
50 m - 100 m	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

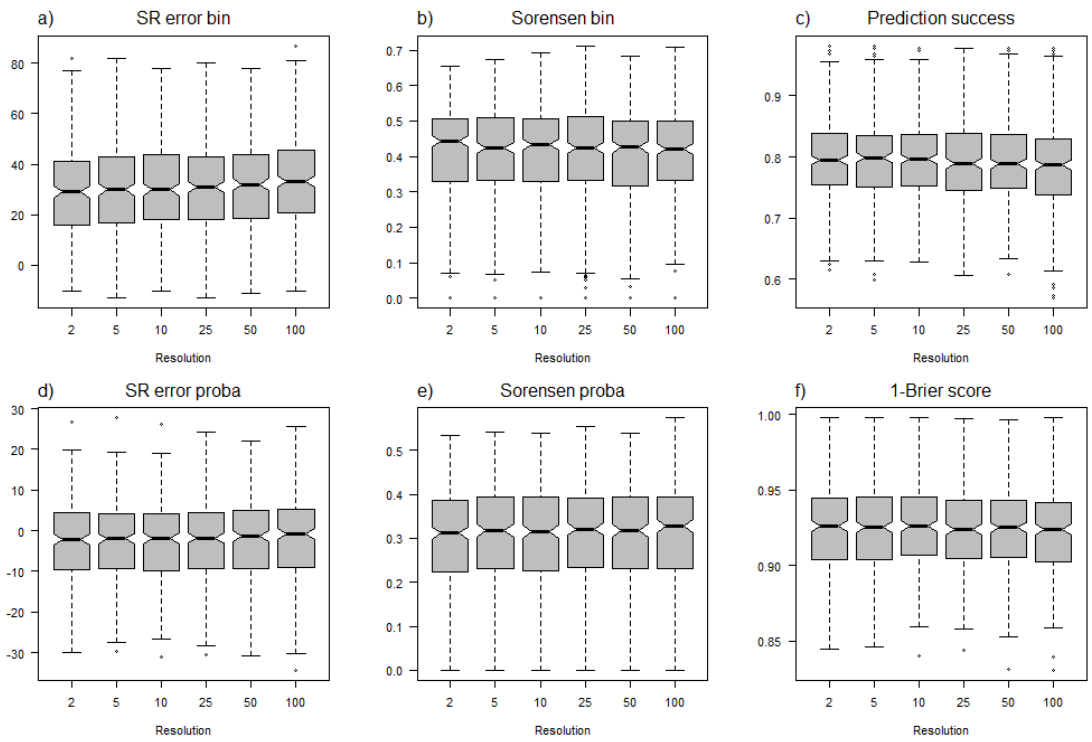


Figure 4. The results of the evaluation of the species assemblage predictions. The first line shows the evaluation of the assemblage prediction after transforming the ensemble forecasting results into binary presences and absences to maximise both sensitivity and specificity: (a) species richness error, (b) Sorensen similarity and (c) prediction success. The second line displays the evaluation results based on comparison of the probabilities predicted by ensemble forecasting with the observed binary data: (d) species richness error, (e) Sorensen similarity and (f) inverse of the Brier score.

Variable importance

Temperature was the most important variable in the SDMs in all cases, followed by the slope, topography and solar radiation (Fig. 5). The differences between the importance of the variables at different resolutions were not significant.

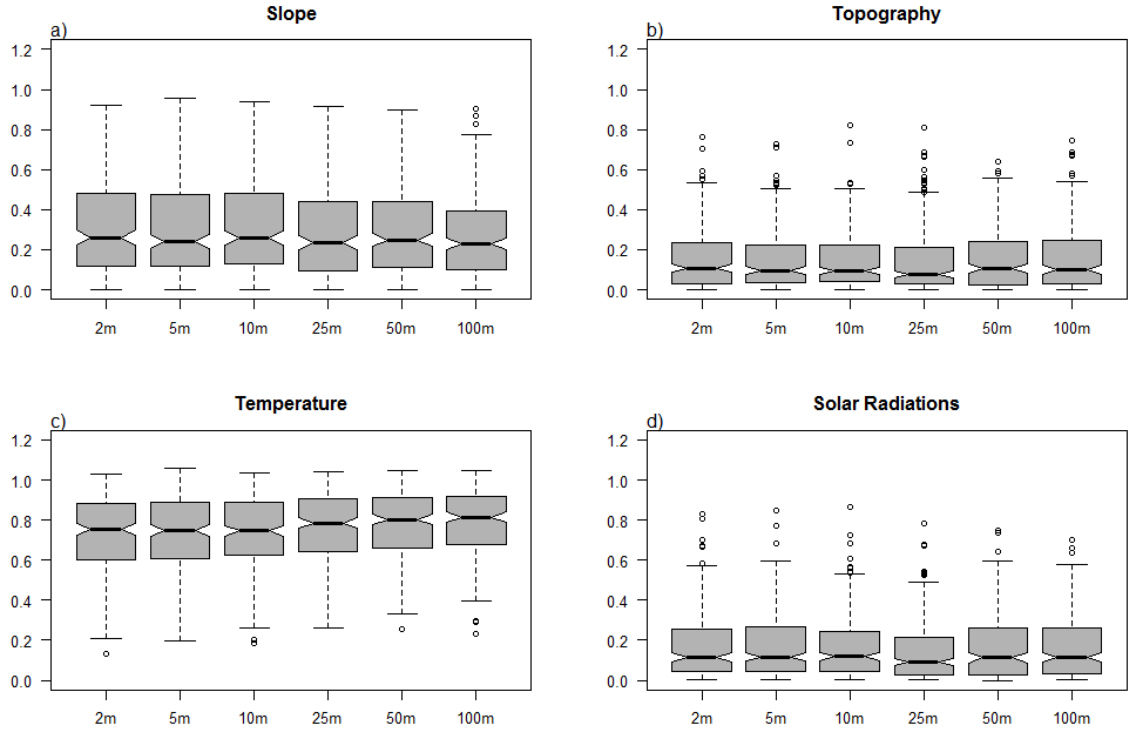


Figure 5. Importance of the variables in the model (mean of the three modelling techniques) for (a) slope, (b) topography,(c) temperature, and (d) solar radiation. The importance of the variables was calculated on the basis of the prediction scores. For each variable, two models were run, one with the unmodified variable and one in which the variable was randomised, also including all other variables in both cases. A correlation score was then calculated between the prediction scores. The importance score for each variable corresponded to one minus the correlation score and could therefore be above one in the case of negative correlations.

Variable correlations

The correlations between the variables at 5 m (the finest resolution, giving the best AUC on average) and 100 m (the coarsest resolution) displayed D^2 values between 0.28 and 0.98 (Fig. 6). The variation of a given variable across the different resolutions depended on the type of variable. The highest variability between the resolutions was observed for the slope and the lowest for the temperature. The variables that were most susceptible to yield variation between models at different resolutions were the slope and topographic position.

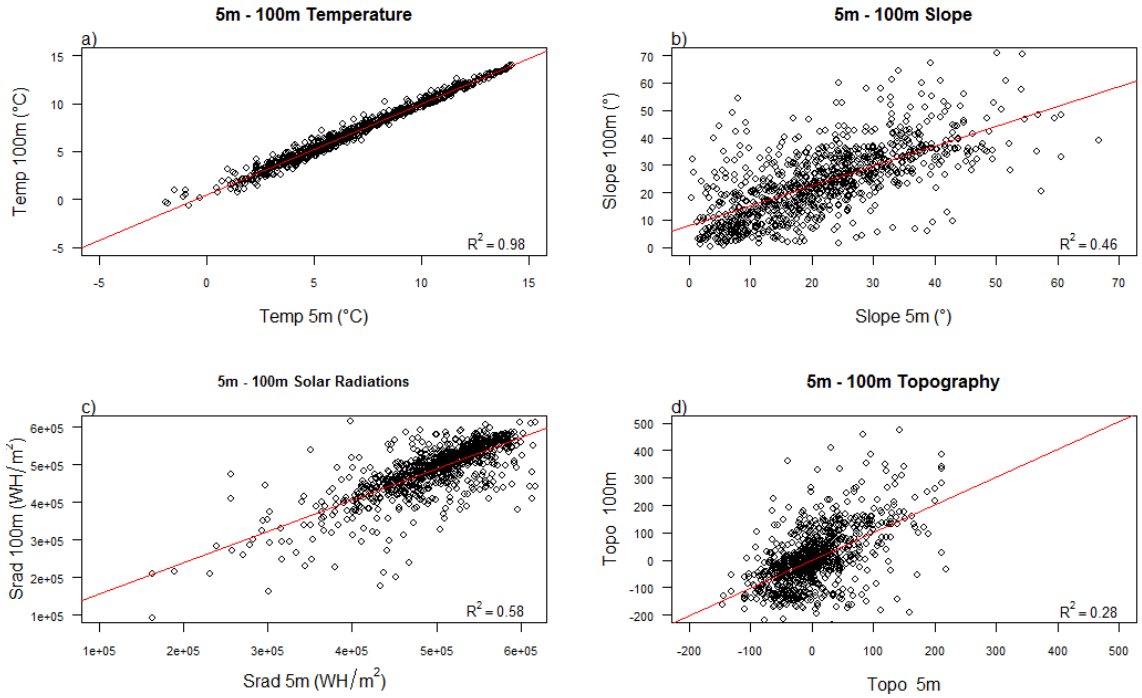


Figure 6. Variable correlations for two resolutions (5 m and 100 m) for (a) temperature, (b) slope, (c) solar radiations and (d) topography. D^2 values are shown in the bottom right corner. The line indicate the trend line between the two resolution.

DISCUSSION

In this study, we compared the performance of plant SDMs based on topographic and temperature predictors at six different resolutions (2, 5, 10, 25, 50, and 100 m). The variations in the predictive accuracy of the models between resolutions were small, showing delta AUC that were rarely greater than ±0.05. Refining the resolution contributed to improving the models slightly for more than half of the species. However, for the other species, the best model remained the one at the coarsest resolution. On average, the best AUC values were obtained for the models at a 5 m resolution, but the differences compared with the other resolutions were small. Five metres may represent a compromise between the finest resolution (2 m), which would cause models to be affected by location errors in the species data sampled in 2 x 2 metre plots, and coarser resolutions (10 m, 25 m, 50 m, and 100 m), which cannot properly represent micro-topographic habitats. The small variations observed in predictive ability can also be explained by the fact that our models showed, on average, consistently good evaluation results across resolutions. Changing the resolution only allowed a refinement of the data and did not add new information. These small variations may also be because temperature, while being the most important variable in the models, was also the predictor that was least variable across resolutions.

Contrary to our initial expectations, we were unable to consistently relate the changes in model performance with the two considered species traits, i.e., the maximal elevation and vegetative height of the species. The overall variation in model performance may remain too small to capture such a relationship, or the chosen traits may not sufficiently reflect the capacity of species to cope with heterogeneous environments, which high-resolution variables are assumed to capture better. Still, for some species known to grow in microhabitats that are likely associated with the fine micro-topography, a particularly large improvement in predictive power was obtained with the refinement of predictor resolution. These species grow, for example, in screes or rocky meadows (*Poa cenisia*, *Senecio doronicum*, *Euphrasia hirtella*, *Thymus spp.*), in areas disturbed by cattle trampling (*Sagina saginoides*, *Poa supina, Plantago major*), or in moist areas *(Carex nigra*, *Viola biflora, Polygonum bistorta*). Nevertheless the increase in predictive power could not be generalised to all species growing in these particular habitats.

Previous studies investigating the impact of changes in grain size on the quality of SDMs have shown divergent results. Guisan *et al.* (2007b) found no differences in predictive power between models for trees built with predictors at 100 m and 1 km. For most of the bird and plant datasets examined by Guisan *et al.* (2007a), coarsening the

predictor resolution ten times caused only a slight decrease of the model's predictive power. Lassueur *et al*. (2006) showed that fine-resolution calculation of the slope increased the predictive power of SDMs when this variable was derived using a neighbourhood window of approximately 100 m. In the present study, however, we did not observe an increasing importance of the slope in the models run at a coarser resolution, which could, for instance, better reflect gravitational processes. More recently, using finer-resolution predictors and bird distribution data, Gottschalk *et al.* (2011) demonstrated a net increase in model quality when improving the resolutions of the predictors from 1 km to 1 m.

The obtained assemblage predictions displayed almost no difference in quality across resolutions. Given the results reported by Seo *et al.* (2009), who showed coarse resolution models to predict larger favourable areas than fine-resolution models, we could have expected the quality of the assemblage predictions to decrease at coarser resolutions. More species would be predicted to be present, causing greater overprediction of species richness and a reduced prediction success. A trend toward an increase in species richness overprediction was observed at the 100 m resolution, but there was no change detected between the models using predictors at 25 m (the resolution used in previous studies) and those at 2 m and 5 m.

Limitations of the study

We mainly used fine numerical elevation data to derive environmental predictors at fine resolutions. However, observational data reflecting the actual micro-conditions experienced by plants at a fine resolution are still lacking, as acquiring these data is difficult, especially in a mountain landscape. Our temperature layers were based on 50 data loggers measuring night-time temperatures. Although night-time temperature has been considered more important than daytime or seasonal mean temperatures (Scherrer & Körner 2011), our fine-resolution temperature maps were not sufficient to improve the SDMs for most species. In this regard, interpolating the night-time temperatures likely caused the resulting maps to be too smoothed to reflect small-scale patterns, such as cold air accumulation or air fluxes between cold and hot air reservoirs, that could change the way the temperature variable affects vegetation in the model (Lundquist, Pepin, & Rochford 2008). Including a greater number of loggers in space to capture these local variations could be a first step forward. A study using VHR predictors coupled with a large number of temperature loggers (e.g., displayed following a stratified design to represent a range of environmental conditions) or with thermal imagery (Scherrer & Körner 2011; Scherrer, Schmid, & Körner 2011) could generate

predictors with a much increased precision. Variables other than temperature like soil moisture or snow cover can take more importance at a micro-scale, but accurate measurements are still lacking for these variables. Direct measurements of predictors in the field as also proved efficient for building SDMs (Le Roux *et al.* 2013). However, applying these new approaches over reasonably large extents, such as our study area (700 km^2 versus 2 km^2 for Scherrer & Körner, 2011 or 0.02 km^2 for Le Roux *et al.* 2013), represents a new challenge going forward for spatial ecologists.

The temperature variations encountered at ground level also depend on the type of substrate involved (Körner 2003). Meadows, bare ground and screes will not exhibit the same temperature patterns and seasonal variations due to differences in heat accumulation and reverberation. Predictor maps of such substrates at a sufficiently fine resolution would also prove useful as direct predictors for plants and for predicting soil properties, but these maps are still rarely available and were not available for our study area. Gathering such data will likely prove important for making progress in future studies. Finally, hyperspectral remote sensing technologies could provide new and complementary approaches to generate high-resolution predictors (Hall *et al.* 2012). Hyperspectral sensors mounted on satellites or airplanes can now gather, in addition to raw bands or vegetation indices (Pouteau *et al.* 2012), data enabling the calculation of temperature or moisture variables at a fine resolution and across large areas (Wang *et al.* 2010).

Conclusions

This study tested the importance of the resolution of the environmental predictors used in species distribution models and the potential improvements conferred by VHR data. We showed that using VHR predictors (i.e., at 2 m, 5 m or 10 m) was slightly beneficial for half of the examined species, while the other species were best predicted at previously applied resolutions (25 m, 50 m, and 100 m). However, no general relationship could be found with plant traits. Our results show that simply increasing the resolution of environmental data by deriving them from a high-resolution digital elevation model is not sufficient to improve models for most species. VHR data are increasingly used but still expensive and difficult to handle over large areas. The costs and the benefits of using these new data in SDMs are not yet trivial. Our results show that monitoring or management studies working with interpolated or modelled data can have similar results working at coarser resolutions (25 to 100 m) but with lower costs for data acquisition and treatment. Our finding suggests that, in future studies, VHR predictors should be developed by also taking into account more local field

measurements, using for instance, climatic micro-loggers and hyperspectral remote sensing technology. These results are important in stressing what the next steps should be for improving predictions of species distributions at a fine scale in complex landscapes, e.g. to derive more realistic scenarios of the impact of climate change on plant distributions (Randin *et al.* 2009a; Scherrer, Schmid, & Körner 2011), to improve the accuracy of species assemblage predictions (Dubuis *et al.* 2011; Pottier *et al.* 2013) at very high elevations, or to identify areas for potential translocation of species following climate change (Thomas 2011; McLane & Aitken 2012).

ACKNOWLEDGMENTS

We thank everyone who helped with data collection. This study was supported by the Swiss National Science Foundation (BIOASSEMBLE project, grant Nr. 31003A-125145 to A. Guisan) and by the European Commission (ECOCHANGE project).

REFERENCES

Allouche, O., Tsoar, A. & 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.

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

Austin, M. (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–19.

Austin, M.P. & Van Niel, K.P. (2011a) Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography*, **38**, 1–8.

Austin, M.P. & Van Niel, K.P. (2011b) Impact of landscape predictors on climate change modelling of species distributions: a case study with Eucalyptus fastigata in southern New South Wales, Australia. *Journal of Biogeography*, **38**, 9–19.

Bertrand, R., Perez, V. & Gégout, J.-C. (2012) Disregarding the edaphic dimension in species distribution models leads to the omission of crucial spatial information under climate change: the case of Quercus pubescens in France. *Global Change Biology*, **18**, 2648–2660.

Camathias, L., Bergamini, A., Küchler, M., Stofer, S. & Baltensweiler, A. (2013) High-resolution remote sensing data improves models of species richness (ed D Rocchini). *Applied Vegetation Science*.

Coudun, C., Gégout, J.-C., Piedallu, C., Rameau, J.-C.C. & Gegout, J.C. (2006) Soil nutritional factors improve models of plant species distribution: an illustration with Acer campestre (L.) in France. *Journal of Biogeography*, **33**, 1750–1763.

Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P. & Guisan, A. (2012) Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables (ed D Rocchini). *Journal of Vegetation Science*, In press.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122–1131.

Ehlers, M., Gaehler, M. & Janowsky, R. (2006) Automated techniques for environmental monitoring and change analyses for ultra high-resolution remote sensing data. *Photogrammetric Engineering & Remote Sensing*, **72**, 835–844.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M. & E. Zimmermann, N. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

Falk, W. & Mellert, K.H. (2011) Species distribution models as a tool for forest management planning under climate change: risk evaluation of Abies alba in Bavaria. *Journal of Vegetation Science*, **22**, 621–634.

Gottfried, M., Pauli, H., Reiter, 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 Distributions*, **5**, 241–251.

Gottschalk, T.K., Aue, B., Hotes, S. & Ekschmitt, K. (2011) Influence of grain size on species–habitat models. *Ecological Modelling*, **222**, 3403–3412.

Guisan, A., Graham, C.H., Elith, J. & Huettmann, F. (2007a) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13**, 332–340.

Guisan, A. & Rahbek, C. (2011) SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.

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.

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? *Ecological Monographs*, **77**, 615–630.

Hall, K., Reitalu, T., Sykes, M.T. & Prentice, H.C. (2012) Spectral heterogeneity of QuickBird satellite data is related to fine-scale plant species spatial turnover in semi-natural grasslands. *Applied Vegetation Science*, **15**, 145–157.

Hanley, J.A. & Mcneil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) Curve. *Radiology*, **143**, 29–36.

Hirzel, A. & Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331–341.

Kalbermatten, M., Van De Ville, D., Turberg, P., Tuia, D. & Joost, S. (2012) Multiscale analysis of geomorphological and geological features in high resolution digital elevation models using the wavelet transform. *Geomorphology*, **138**, 352–363.

Körner, C. (2003) *Alpine Plant Life (2nd Edition)*. Springer, New York.

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

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

Lundquist, J.D., Pepin, N. & Rochford, C. (2008) Automated algorithm for mapping regions of cold-air pooling in complex terrain. *Journal of Geophysical Research*, **113**, D22107.

McGill, B.J. (2010) Matters of scale. *Science*, **328**, 575–6.

McLane, S.C. & Aitken, S.N. (2012) Whitebark pine (Pinus albicaulis) assisted migration potential: testing establishment north of the species range. *Ecological Applications*, **22**, 142–153.

Nogués-Bravo, D. & Rahbek, C. (2011) Ecology. Communities under climate change. *Science*, **334**, 1070–1.

Pellissier, L., Bråthen, K.A., Vittoz, P., Yoccoz, N.G., Dubuis, A., Meier, E.S., Zimmermann, N.E., Randin, C.F., Thuiller, W., Garraud, L., Van Es, J. & Guisan, A. (2013) Thermal niches are more conserved at cold than warm limits in arctic-alpine plant species. *Global Ecology and Biogeography*, **in press**.

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P. & Guisan, A. (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients (ed R Field). *Global Ecology and Biogeography*, **22**, 52–63.

Pouteau, R., Meyer, J.-Y., Taputuarai, R. & Stoll, B. (2012) Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecological Informatics*, **9**, 37–46.

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

Randin, C.F., Jaccard, H., Vittoz, P., Yoccoz, N.G. & Guisan, A. (2009b) Land use improves spatial predictions of mountain plant abundance but not presence-absence. *Journal of Vegetation Science*, **20**, 996–1008.

Randin, C.F., Vuissoz, G., Liston, G.E., Vittoz, P. & Guisan, A. (2009c) Introduction of Snow and Geomorphic Disturbance Variables into Predictive Models of Alpine Plant Distribution in the Western Swiss Alps. *Arctic, Antarctic, and Alpine Research*, **41**, 347–361.

Le Roux, P.C., Lenoir, J., Pellissier, L., Wisz, M.S. & Luoto, M. (2013) Horizontal, but not vertical, biotic interactions affect fine-scale plant distribution patterns in a low energy system. *Ecology*.

Le Roux, P.C., Virtanen, R. & Luoto, M. (2013) Geomorphological disturbance is necessary for predicting fine-scale species distributions. *Ecography*.

Scherrer, D. & Körner, C. (2011) Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *Journal of Biogeography*, **38**, 406–416.

Scherrer, D., Schmid, S. & Körner, C. (2011) Elevational species shifts in a warmer climate are overestimated when based on weather station data. *International journal of biometeorology*, **55**, 645–54.

Seo, C., Thorne, J.H., Hannah, L. & Thuiller, W. (2009) Scale effects in species distribution models: implications for conservation planning under climate change. *Biology letters*, **5**, 39–43.

Thomas, C.D. (2011) Translocation of species, climate change, and the end of trying to recreate past ecological communities. *Trends in ecology & evolution*, **26**, 216–21.

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

Vaze, J., Teng, J. & Spencer, G. (2010) Impact of DEM accuracy and resolution on topographic indices. *Environmental Modelling & Software*, **25**, 1086–1098.

Wang, K., Franklin, S.E., Guo, X. & Cattet, M. (2010) Remote sensing of ecology, biodiversity and conservation: a review from the perspective of remote sensing specialists. *Sensors (Basel, Switzerland)*, **10**, 9647–67.

Wang, W., Lo, N., Chang, W. & Huang, K. (2012) Modeling spatial distribution of a rare and endangered plant species (Brainea insignis) in central Taiwan. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **XXXIX-B7**, 241–246.

Woodward, F.I. (1987) *Climate and Plant Distribution*. Cambridge University Press.

SUPPLEMENTARY INFORMATIONS

Table S1. Evaluation of temperature models at different grid sizes for June, July and August based on the D^2, adjusted D^2, and root mean square error (RMSE) computed after cross-validation.

		June	July	August
2 m	**D2**	0.770	0.781	0.769
	adj D2	0.755	0.766	0.753
	RMSE	1.364	1.374	1.398
5 m	**D2**	0.760	0.772	0.758
	adj D2	0.744	0.756	0.741
	RMSE	1.665	1.665	1.709
10 m	**D2**	0.764	0.774	0.761
	adj D2	0.748	0.759	0.745
	RMSE	1.430	1.431	1.460
25 m	**D2**	0.766	0.778	0.767
	adj D2	0.750	0.763	0.751
	RMSE	1.395	1.400	1.420
50 m	**D2**	0.787	0.796	0.761
	adj D2	0.773	0.782	0.746
	RMSE	1.324	1.341	1.421
100 m	**D2**	0.761	0.768	0.752
	adj D2	0.746	0.753	0.736
	RMSE	1.421	1.452	1.498

Table S2. The means, variances and standard deviations for the AUC values for the three modelling techniques.

	2m			5m			10m			25m			50m			100m		
	GAM	GBM	GLM	GAM	GBM	GLM	GAM	GBM	GLM	GAM	GBM	GLM	GAM	GBM	GLM	GAM	GBM	GLM
Mean	0.807	0.776	0.807	0.809	0.780	0.809	0.807	0.780	0.805	0.802	0.778	0.801	0.802	0.782	0.799	0.794	0.774	0.794
Variance	0.006	0.007	0.005	0.006	0.007	0.005	0.006	0.007	0.006	0.006	0.007	0.006	0.006	0.007	0.006	0.007	0.007	0.006
SD	0.075	0.085	0.073	0.075	0.081	0.070	0.075	0.082	0.074	0.079	0.083	0.078	0.076	0.084	0.075	0.081	0.081	0.080
	Techniques Averaged			Techniques Averaged			Techniques Averaged			Techniques Averaged			Techniques Averaged			Techniques Averaged		
Mean		0.797			0.799			0.797			0.794			0.794			0.787	
Variance		0.006			0.006			0.006			0.006			0.006			0.006	
SD		0.078			0.075			0.077			0.080			0.078			0.081	

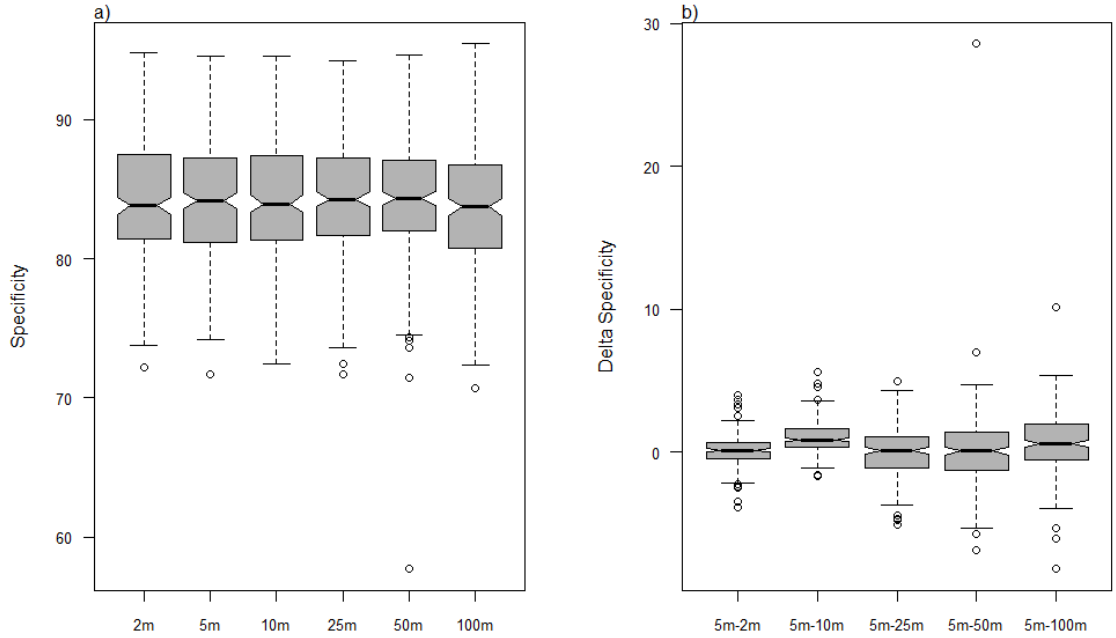


Figure S1. (a) Boxplot of the specificity values calculated from the AUC (mean of the three modelling techniques for each species) according to the different resolutions. (b) Comparison of the specificity of the 5 m models (with the specificity of those run at 10, 25, 50 and 100 m resolutions).

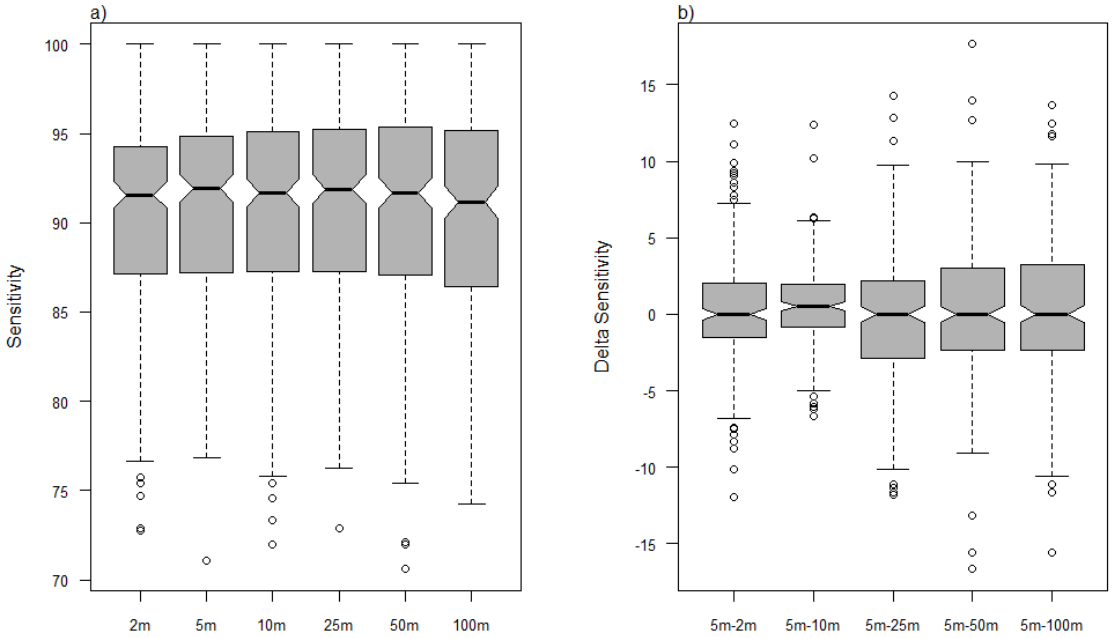


Figure S2. (a) Boxplot of the sensitivity values calculated from the AUC (mean of the three modelling techniques for each species) according to the different resolutions. (b) Comparison of the sensitivity of the 5 m models (with the sensitivity of those run at 10, 25, 50 and 100 m resolutions).

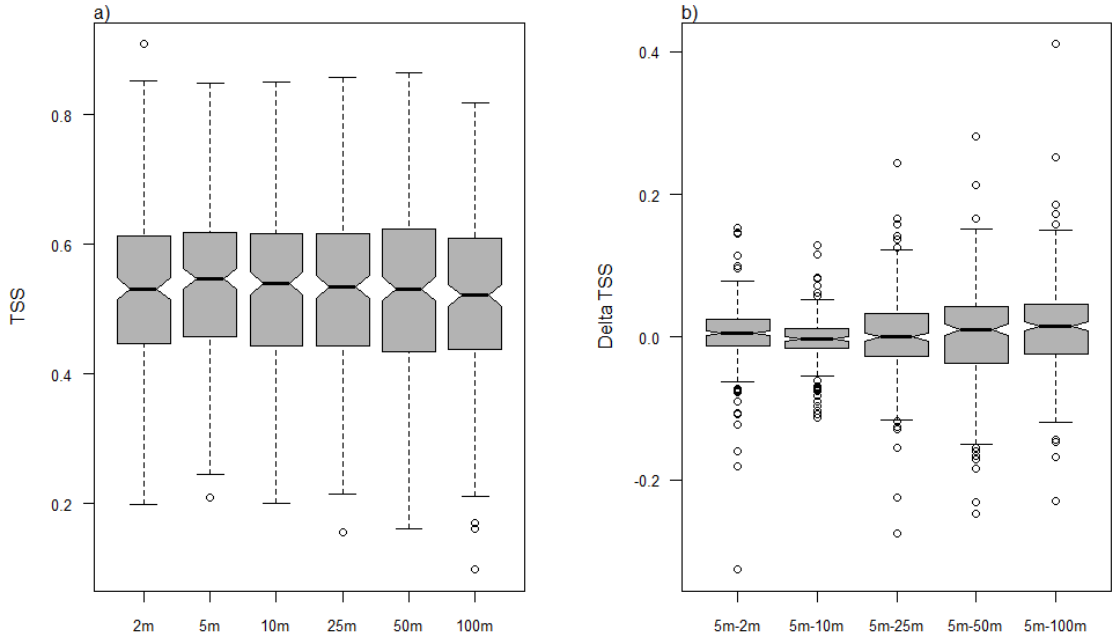


Figure S3. (a) Boxplot of the TSS values (mean of the three modelling techniques for each species) according to the different resolutions. (b) Comparison of the TSS for the 5 m models (with the TSS for those run at 10, 25, 50 and 100 m resolutions).

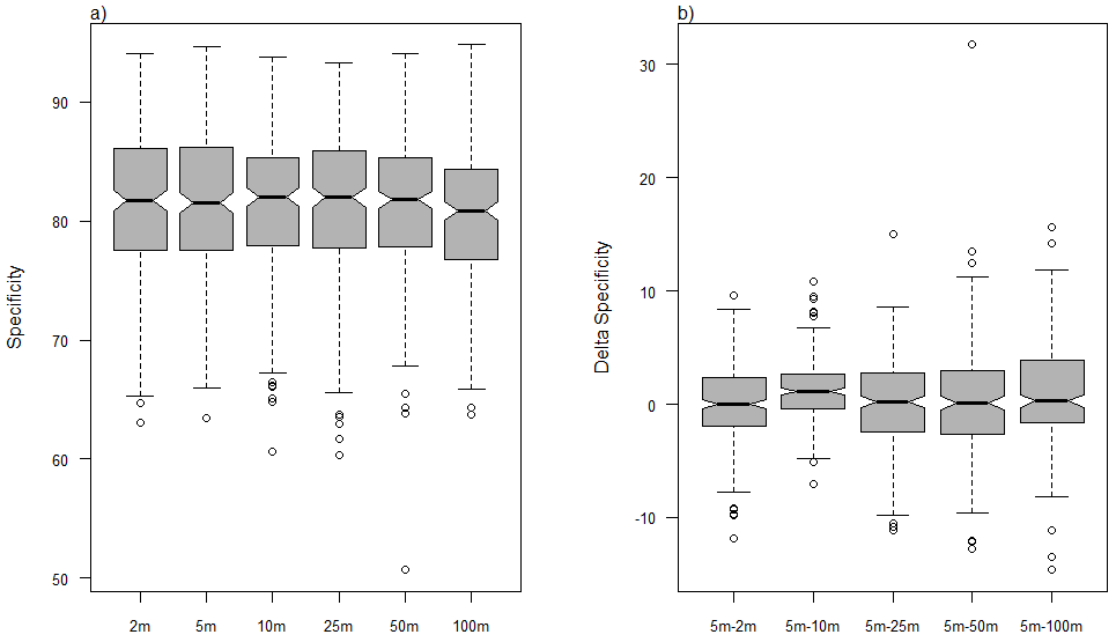


Figure S4. (a) Boxplot of the specificity values calculated from the TSS (mean of the three modelling techniques for each species) according to the different resolutions. (b) Comparison of the specificity of the 5 m models (with the specificity of those run at 10, 25, 50 and 100 m resolutions).

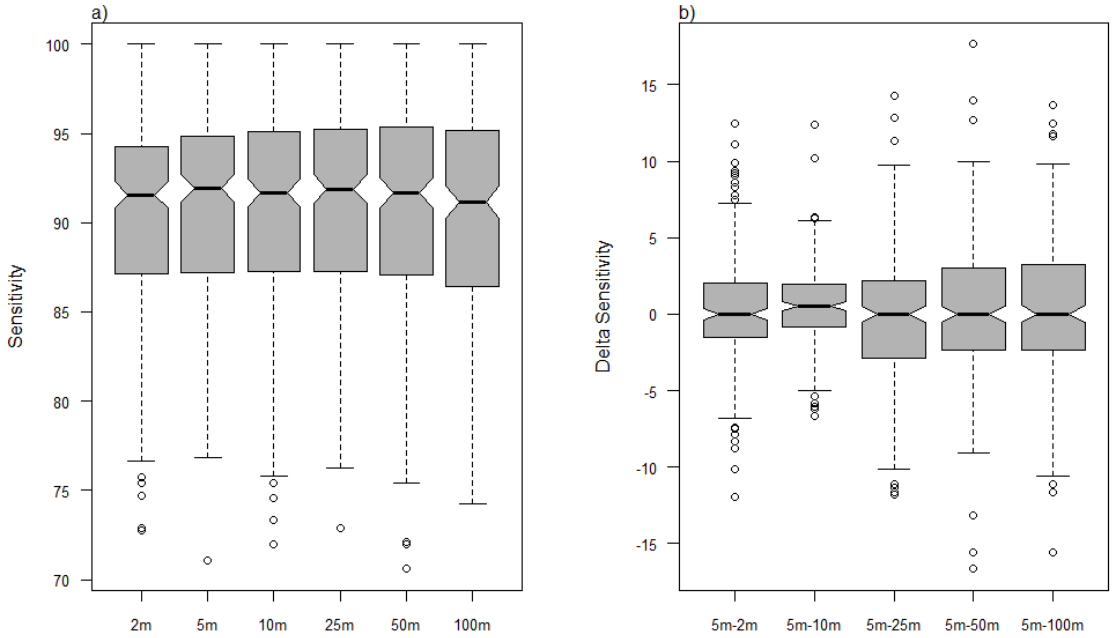


Figure S5. (a) Boxplot of the sensitivity values calculated from the TSS (mean of the three modelling techniques for each species) according to the different resolutions. (b) Comparison of the sensitivity of the 5 m models (with the sensitivity of those run at 10, 25, 50 and 100 m resolutions).

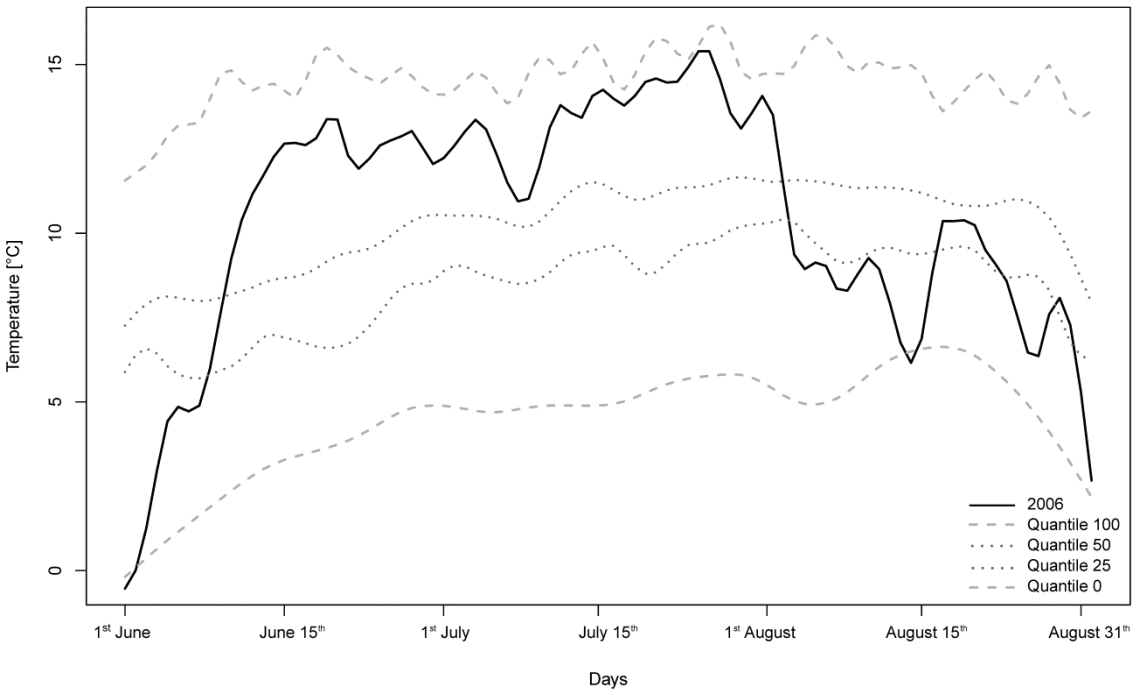


Figure S6. Daily average of temperature for the months of June, July and August (black line) calculated from the five MeteoSuisse stations in the study area and the corresponding quantiles (dotted lines = 25% and 50%, dashed lines = 0% and 100%) for the hindcast period (1981-2010) for these five stations.

CHAPTER 3

Predicting spatial patterns of plant species richness: a comparison of direct macro-ecological and species stacking modelling approaches

Anne Dubuis [1]
Julien Pottier [1]
Vanessa Rion [1]
Loïc Pellissier [1]
Jean-Paul Theurillat [3, 4]
Antoine Guisan [1]

1 Department of Ecology and Evolution, University of Lausanne, Bâtiment Biophore, 1015 Lausanne, Switzerland;

2 Fondation J.-M. Aubert, 1938 Champex-Lac, Switzerland

3 Laboratory of Biogeography, Section of Biology, University of Geneva, 1292 Chambésy, Switzerland

This paper has been published in 2011
Diversity and Distributions **17**, 1122-1131
DOI: 10.1111/j.1472-4642.2011.00792.x

Contribution to the project: I participated to the field work for plant data collection. I performed the species distribution modelling and analysed the results. I wrote the initial draft of the paper and lead the reviewing process.

ABSTRACT

Aim: This study compares the direct, macroecological approach (MEM) for modelling species richness with the more recent approach of stacking predictions from individual species distributions (S-SDM). We implemented both approaches on the same dataset, and discuss their respective theoretical assumptions, strengths and drawbacks. We also tested how both approaches performed in reproducing observed patterns of species richness along an elevation gradient.

Location: Two study areas in the Alps of Switzerland.

Methods: We implemented MEM by relating the species counts to environmental predictors with statistical models, assuming a Poisson distribution. S-SDM was implemented by modelling each species distribution individually and then stacking the obtained prediction maps in three different ways – summing binary predictions, summing random draws of binomial trials and summing predicted probabilities – to obtain a final species count.

Results: The direct MEM approach yields nearly unbiased predictions centred around the observed mean values, but with a lower correlation between predictions and observations, than that achieved by the S-SDM approaches. This method also cannot provide any information on species identity and, thus, community composition. It does, however, accurately reproduce the hump-shaped pattern of species richness observed along the elevation gradient. The S-SDM approach summing binary maps can predict individual species and thus communities, but tends to overpredict species richness. The two other S-SDM approaches - the summed binomial trials based on predicted probabilities and summed predicted probabilities - do not overpredict richness, but they predict many competing end points of assembly or they lose the individual species' predictions, respectively. Furthermore, all S-SDM approaches fail to appropriately reproduce the observed hump-shaped patterns of species richness along the elevation gradient.

Main conclusions: MEM and S-SDM have complementary strengths. We suggest that both could be used in combination to obtain better species richness predictions by following the suggestion of constraining S-SDM by MEM predictions.

Keywords

macroecological models, plants, species richness, stacked species distribution models, Swiss Alps

INTRODUCTION

With the increasing human impact on biodiversity (Sala *et al.*, 2000), the factors that determine the distributions of species and the communities they compose are of primary interest to the optimization of conservation actions. Species richness (SR), that is, the number of taxa occurring in a defined geographic unit, is widely used as a measure of biodiversity (Whittaker, 1972). However, information on SR is often incomplete, with large areas being devoid of such data. Empirical modelling of SR can be used to overcome this limitation, for example by identifying hotspots of SR (Gioia & Pigott, 2000; Lehmann *et al.*, 2002; Luoto *et al.*, 2004; Parviainen *et al.*, 2009). This information can then be used to help establish conservation strategies or to predict future patterns of biodiversity under global change (Guisan & Theurillat, 2000a; Algar *et al.*, 2009). These models can also help to improve our understanding of biodiversity patterns and community assembly processes (Gioia & Pigott, 2000; Gould, 2000; Moser *et al.*, 2005; Thuiller *et al.*, 2006; Nogues-Bravo *et al.*, 2008).

Two different approaches have been used to model SR: direct modelling of species numbers, or stacking of individual species predictions (Ferrier & Guisan, 2006). The first approach, hereafter referred to as “macroecological modelling” (MEM), statistically relates SR – the count of species within a geographic unit – to values of environmental variables that characterize the same unit (Gould, 2000; Luoto *et al.*, 2004; Moser *et al.*, 2005). This approach has been used to better understand (Rahbek & Graves, 2001; Jetz & Rahbek, 2002; Moser *et al.*, 2005) and predict (Luoto *et al.*, 2004) SR for numerous taxonomic groups at a wide variety of scales. In this macroecological approach, the number of species found at a given location is expected to depend on various control factors such as unit size, resources available (Pausas & Austin, 2001; Hawkins *et al.*, 2003; Michalet *et al.*, 2006; Whittaker *et al.*, 2007), environmental heterogeneity (Palmer & Dixon, 1990; Dufour *et al.*, 2006) and disturbance levels (Connell, 1978; Huston, 1979; Mackey & Currie, 2000, 2001).

The second approach, stacked species distribution modelling (S-SDM, Guisan & Rahbek, in press), consists of first predicting the distribution of each species independently using species distribution models (SDM; Guisan & Thuiller, 2005) and then stacking them to predict species assemblages, providing both SR and composition for each modelled unit. S-SDMs have already been used to predict current (Cumming, 2000; Parviainen *et al.*, 2009; Pineda & Lobo, 2009) and future (Guisan & Theurillat, 2000b; Algar *et al.*, 2009; Pineda & Lobo, 2009) distributions of SR, community composition (Pineda & Lobo, 2009), and species turnover rates

(Thuiller *et al.*, 2005). SDMs statistically relate species occurrences to a set of environmental variables (Guisan & Zimmermann, 2000), thus relying on the realized environmental niche concept introduced by Hutchinson (1957), including biotic and dispersal limitations (Pulliam, 2000; Soberon, 2007). One could then assume that combining individual species predictions allows the prediction of SR at each modelled unit. The theoretical assumption underpinning the method of stacking individual SDMs (hereafter S-SDMs), follows from the ideas of Gleason (1926) and, more recently, Ricklefs (2004, 2008), stating that a local assemblage is made of a set of species with partially the same environmental requirements (habitat filtering, Guisan & Rahbek, in press)

How different are these approaches? Although both depend on the the availability of a regional species pool (i.e. taking into account biogeographic history; Aarssen & Schamp, 2002), they differ markedly in many other theoretical and methodological aspects (Guisan & Rahbek, in press). From the theoretical perspective, MEM relies on the existence of macroecological controls on community assembly whereas S-SDM relies on a Gleasonian overlay of species and therefore inherits assumptions typically associated with SDM, such as equilibrium and niche stability (see Guisan & Rahbek (in press) for more detailed explanations). This means that even using the same set of predictors, the underlying assumptions behind MEM and S-SDM differ, with the same factors possibly translating different causal processes. For instance, degree-days might mostly express the amount of energy that can be shared between species in MEM, but would rather express the limit to the growth period of single species in SDMs.

From the methodological point of view, MEMs, as applied in the vast majority of published studies, are fitted using standard statistical methods usually involving some form of averaging. As such, they should predict mean SR with symmetrically distributed errors. A direct corollary is that the most extreme SR values may be more difficult to predict (i.e. average methods tend to underpredict the highest values and overpredict the lowest values). However, the main limitation of this approach remains that it does not provide information regarding the identity of the species occurring at a given location and thus cannot predict changes in species composition.

On the other hand, S-SDMs can predict species composition but do not include a way to set a limit on the number of species that can possibly occupy a given habitat. S-SDMs thus ignore the environmental controls on SR that are hypothesized by MEMs, especially in saturated communities with productive habitats and intense competition

for resources. In such communities, the set of species actually occurring at a given site is sorted from a larger habitat species pool, as determined by neutral processes and/or biotic assembly rules (Cornell & Lawton, 1992; Loreau, 2000). Without rules to constrain SR, S-SDMs are expected to predict the entire pool of suitable species for a habitat, potentially overpredicting SR. Moreover, the way in which SDMs with probabilistic predictions are often transformed into binary presence-absence predictions is subject to discussion and could also be causing overprediction of SR. For instance, Pineda & Lobo (2009) initially selected a single threshold applied across all species and then adapted this threshold for individual species that were over- or underestimated. They showed that this method allowed for correction of the predicted SR overestimation, but the rationale for correcting thresholds lacks support from ecological theory. Another approach is to sum the predicted probability directly (Lehmann *et al.*, 2002; Gelfand *et al.*, 2005). Hence, there are different ways in S-SDMs of stacking species' predictions to assemble communities, but still few studies have evaluated them. Finally, another limitation of S-SDMs is that a sufficient number of occurrences must be available for each species to be modelled. Species that are too rare may need to be excluded from the analyses, whereas they can be implicitly included in the MEM.

Given these different theoretical assumptions and methodological implications, MEMs and S-SDMs may therefore be expected to yield different SR predictions. How then do they perform relative to each other in predicting plant SR? Are they equally successful or does one approach out-perform the other? Answers to these questions are still scarce, as very few studies have compared the two approaches comprehensively using a same dataset. Lehmann et al. (2002) compared MEM and S-SDM for predicting ferns SR in New-Zealand, Algar et al. (2009) compared them for butterflies in Canada, additionally testing their predictive power across time during the 20th century, and Newbold et al. (2009) used both approaches to map mammal and butterfly diversity patterns in Egypt. However, none of these studies really discuss their results in light of the theoretical expectations of these approaches, and especially the expectation that S-SDM should tend to overpredict SR compared to MEM.

Here we use a comprehensive dataset of vegetation plots sampled across two distinct mountain study areas in the Swiss Alps (Guisan *et al.*, 1998; Guisan & Theurillat, 2000a; Randin *et al.*, 2006; Engler *et al.*, 2009; Randin *et al.*, 2009b) and topo-climatic predictors to implement and compare the MEM and three different S-SDM approaches using binarized predictions, random binomial realizations and summed probabilities,

respectively. For each modelling approach, we considered both (i) its accuracy in predicting SR and (ii) its ability to reproduce actual SR patterns observed in relation to elevation, the most important environmental gradient in both study areas. As a null hypothesis, we consider that the values of SR predicted by MEMs and S-SDMs do not differ. As an alternative hypothesis, we hypothesize that MEM will predict richness values around the observed mean, whereas S-SDM will predict an excessive number of species. We further expect that the magnitude of overprediction depends on how single predictions are stacked.

METHODS

Study areas

The Diablerets study area, covering ca. 700 km^2, is located in the western Swiss Alps (Vaud, Switzerland, 46°10' to 46°30'N; 6°60' to 7°10'E), with elevations ranging between 375 m and 3,210 m. The climate is temperate: annual temperatures and precipitation fluctuate from 8 °C and 1,200 mm at 600 m to -5 °C and 2,600 mm at 3,000 m (Bouët, 1985). The soil parent material is mainly calcareous. Human influence on vegetation in this area is extensive. Non-forested areas are often used as pastures or are mowed, and fertilization is important from the lowlands to the subalpine belt. For more information on this area, see Randin *et al.*(2006).

The Belalp study area, covering ca. 20 km^2, is located in the Aletsch region (Valais, Switzerland, 46°19 to 46°24'N; 7°55' to 7°59'E). This area is a north-south oriented side valley of the Rhone valley. Elevation ranges from 1,867 m to 3,554 m. The climate is subcontinental. The soil parent material is siliceous (Steck, 1983). Human influence mainly consists of livestock grazing. For more information on this area, see Guisan *et al.* (1998).

Vegetation data

The vegetation data used in our analysis originate from different field surveys conducted between 1993 and 1995 in Belalp and between 2002 and 2009 in the Diablerets. In Belalp, 205 plots of 4 m^2 were inventoried following a grid-sampling scheme. In the Diablerets, 912 4 m^2 vegetation plots were inventoried according to a random-stratified sampling design (Hirzel & Guisan, 2002) based on elevation, slope and aspect. The sampling was limited to vascular species in open and non-woody vegetation only, i.e., grasslands, meadows, rocks and screes. Each selected point was separated from the others by a minimum distance of 200 m to avoid spatial autocorrelation. Only species with more than 20 occurrences throughout all sampled plots were kept for further analyses. This approach resulted in 260 species in the Diablerets and 42 in Belalp.

Climatic and topographic predictors

For the Diablerets area, we used three bioclimatic predictors that have been demonstrated to have a direct ecophysiological impact on plant species (Körner, 1999; Dirnböck *et al.*, 2003). We used two topographic predictors that were assumed

to account for gravitational processes such as snow avalanches, rockfalls and drainage effects (Guisan *et al.*, 1998; Dirnböck *et al.*, 2003). Degree-days (with 0°C threshold), moisture index over the growing season (difference between precipitation and potential evapotranspiration from June to August) and global solar radiation during the growing season were derived at a resolution of 25 m from temperature and precipitation values interpolated from a network of meteorological stations and from GIS-derived solar radiation. Slope and topographic position were derived from the digital elevation model. The methods used to compute these predictors in the Diablerets area are fully described in Randin *et al.* (2006).

For the Belalp area, three bioclimatic predictors - annual mean temperature, solar radiation and a moisture index - were also considered, and the same two topographic predictors - slope and topographic position - were derived and used at the 25 m resolution following similar procedures. The methods used to compute environmental predictors in the Belalp area are fully described in Guisan *et al.* (1998).

These variables have already been used successfully in several modelling studies in the same areas (Guisan *et al.*, 1998; Guisan & Theurillat, 2000a; Guisan & Thuiller, 2005; Randin *et al.*, 2006; Engler *et al.*, 2009; Randin *et al.*, 2009b). We did not use landuse or geological predictors for two reasons: because (i) they are not available in a fully spatially-explicit way for the study area; and (ii) two previous studies testing their predictive power for plants in subareas revealed effects largely nested within those of topoclimate (Randin *et al.*, 2009b; Randin *et al.*, 2009a), suggesting that the above-mentioned topo-climatic predictors are enough to predict the presence-absence of most species.

These environmental data are known to correlate well with vegetation data, due to the accuracy with which plots are located in the field (< 1 m for the Diablerets, ~5 m for Belalp), being much finer than the 25 m resolution of the environmental maps, and due to the high accuracy of the Swiss 25 m DEM. Finally, all environmental predictors had pairwise correlations < 0.8, limiting the risk of multi-colinearity in the models.

Approaches to modelling species richness

For the sake of comparison between MEM and S-SDM, we used the same four modelling techniques in both cases. These techniques are representative of current common practice. Due to several SDM techniques being inadequate for fitting MEM, we used only those available for both approaches. We implemented, as much as

possible, the same parameterization of each technique for the two modelling approaches.

Macroecological models

In the direct approach, SR is set as the response variable in statistical models using a Poisson distribution (Fisher *et al.*, 1943; Vincent & Haworth, 1983). Here, SR was calculated as the count of all species also modelled with SDMs (i.e., occurring in more than 20 plots; 260 in the Diablerets and 42 in Belalp).

Four modelling techniques were used: generalized linear models (GLM; McCullagh & Nelder, 1989; R library "glm"); generalized additive models (GAM; Hastie & Tibshirani, 1990; R library "gam"); generalized boosted models (GBM; Ridgeway, 1999; Friedman *et al.*, 2000; Friedman, 2001; R library "gbm"); and random forests (RF; Breiman, 2001; R library "randomForest"). GLM was implemented using a quadratic function and GAM with smoothing functions with up to four degrees of freedom, and both were calibrated using a Poisson distribution and a logarithmic link function (i.e., logistic regression). GBM was implemented using a maximum of 5,000 trees, and a Poisson distribution. Random Forest was also implemented using 5,000 trees.

The accuracy of each model was evaluated with a repeated (100 times) split-sample procedure. An evaluation dataset was obtained by randomly splitting the original dataset into two 70%-30% partitions, using the 70% partition for fitting the models and the other 30% for independently evaluating these models. For each split-sample repetition and for each model, a Spearman rank correlation between observed and predicted SR was calculated using the evaluation data set. The final SR predictions were obtained by averaging the predictions from the four different techniques, and these were again evaluated with Spearman correlations.

Stacked species distribution models

In the S-SDM, each species was first modelled separately. The BIOMOD library in R (Thuiller *et al.*, 2009) was used to implement GLM, GAM, GBM and RF, assuming a binomial distribution and the same parameterizing options as for the MEM, but separately for each species (260 in the Diablerets and 42 in Belalp). Again, we used a repeated (100 times) split-sample approach for evaluating models. Each model was fitted using 70% of the plots and evaluated using the area under the curve (AUC) of a receiver-operating characteristics (ROC) plot (Fielding & Bell, 1997) calculated on the excluded 30%.

The projected distributions for individual species were then stacked in three different ways. In the first approach, the maps were reclassified into presence and absence using a ROC-optimized threshold, as has often been described previously in the literature (Fielding & Bell, 1997; Liu *et al.*, 2005; see also Algar *et al.*, 2009), thus maximizing model sensitivity and specificity. These presence-absence layers were summed to obtain an SR map for every modelling technique. The average of these four maps produced the final SR map ($S\text{-}SDM_{binroc}$). To avoid choosing any threshold, which can affect the predictions, we produced a second set of richness maps directly from raw probabilities predicted by the SDMs (averaged across the four modeling techniques). To do so, we first produced a large number of richness maps (i.e., 1,000), each being the result of stacking presence/absence maps of species randomly generated from binomial trials using their respective predicted probability as a parameter. We then averaged SR over the repeated trials ($S\text{-}SDM_{resamp}$). We also produced a richness map by simply summing the mean probabilities (Lehmann *et al.*, 2002; Gelfand *et al.*, 2005) predicted by each technique for each species in each plot ($S\text{-}SDM_{sumprob}$). Note that the last two approaches should theoretically produce the same results, with a difference tending toward zero as the trial number in the $S\text{-}SDM_{resamp}$ increases toward infinity. The predictive accuracy of the three final S-SDM maps was again evaluated using a Spearman correlation between observed and predicted SR.

Comparison of the two approaches

To further evaluate the accuracy of predictions, the richness predicted by different modelling approaches was plotted against the observed richness, and a linear regression line was fitted to this plot. A good predictive model is expected to have an intercept close to zero and a slope close to one (Algar *et al.*, 2009). To assess how the modelling approaches performed in reproducing well-known patterns of species diversity, we fitted regressions with quadratic terms to analyze the relationship between SR predictions and elevation, and we compared these relationships with that obtained using observed SR against elevation.

Model calibration and other statistical analyses were computed using the R software version 2.10.0 (R Development Core Team, 2009).

RESULTS

Predictions of species richness (SR) were consistent across the four modelling techniques (i.e., GLM, GAM, GBM and RF). Therefore, in the following sections we present only results based on the ensemble predictions averaged across techniques. Furthermore, the results for Belalp were globally consistent with those for the Diablerets. However, the Belalp area encompasses a much smaller elevation range than the Diablerets and, therefore, compares only with the upper elevations of the Diablerets area, where smaller differences between the MEM and S-SDM approaches were observed. Therefore, we here present results only for Diablerets. All results for Belalp are presented in Figures S1, S2 and S3 of Supporting Information.

Macroecological models

The MEM-predicted SR achieved a Spearman rank correlation of 0.67 with observed SR (Figure 1a). Plots with the lowest observed SR tended to be overestimated, whereas plots with SR > 30 tended to be underestimated. This effect was best captured by the regression line drawn through the scatter plot of the observed versus predicted SR, with an intercept of 13.86 and a slope of 0.42. The histogram of prediction errors (Figure 1d) revealed that the proportion of over- and underpredicted plots were quite similar, resulting in an overall prediction bias of nearly zero (mean error = -0.08).

Stacked species distribution models

AUCs of all models for both areas ranged between 0.5 (poor) and 0.99 (excellent) (Figure S5), but the vast majority (92%) were over 0.7 (useful models according to Swets (1988)). SR predicted by S-SDM$_{binroc}$ was correlated with a Spearman rho of 0.68 with the observed SR (Figure 1b). The regression line through the scatter plot had an intercept of 37.33 and a slope of 0.83, revealing strong and constant overpredictions of SR, additionally shown by the histogram of model errors (Figure 1e). Here, unlike for MEMs (Fig. 1d), the prediction bias was high and constant (mean error = 33.17). The richness map generated from binomial trials (S-SDM$_{resamp}$) was correlated with a rho of 0.79 with the observed data (Figure 1c). The overprediction became, on average, negligible (mean error of 0.48, Figure 1f): species-poor plots are overestimated (intercept of 14.26) while species-rich plots are underestimated (slope of 0.42). This approach allows an estimation of uncertainty, which, in this case, is the standard deviation of the 1,000

resampling ranges from 3.4 to 5.7 species among the plots. Finally, the summed probabilities approach (S-SDM$_{sumprob}$) generated results nearly identical to those of S-SDM$_{resamp}$. These results are presented in figure S4.

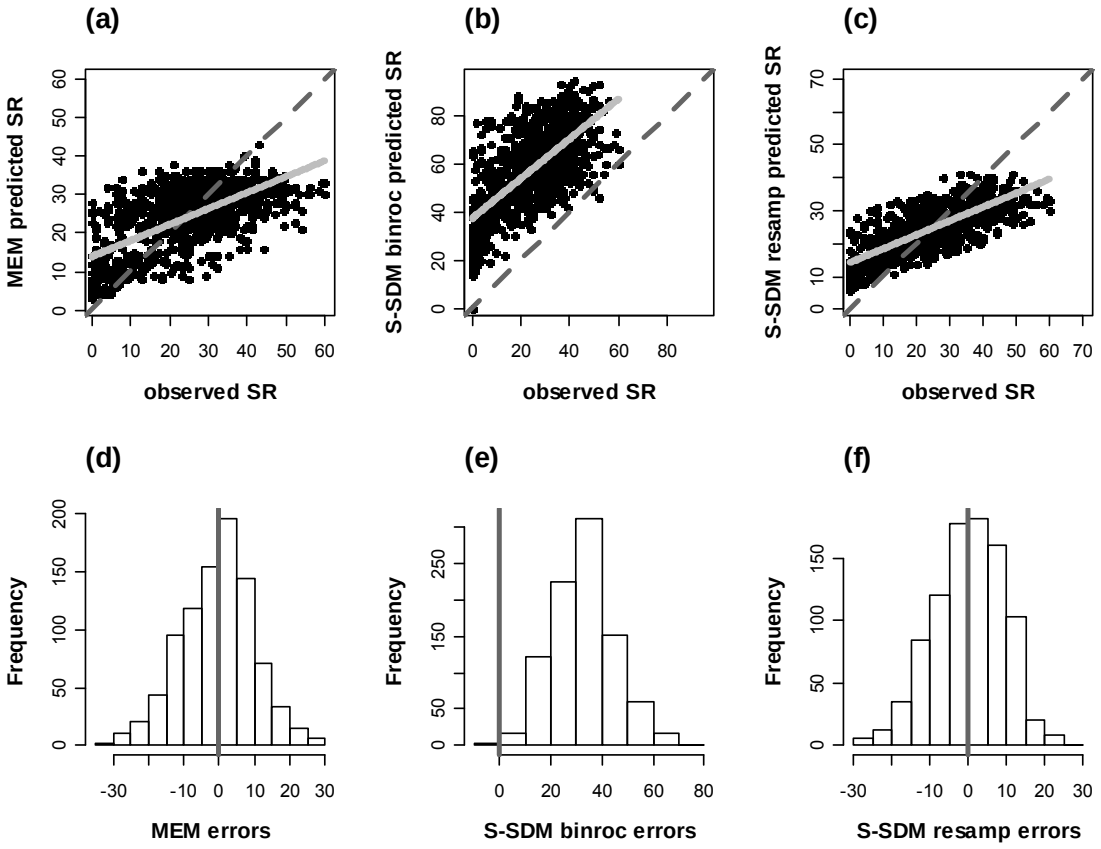


Figure 1: First line: predicted SR plotted against the observed data for (a) MEM, (b) S-SDM$_{binroc}$, and (c) S-SDM$_{resamp}$. The dashed line shows a perfect correspondence between observed and predicted SR, and the plain line corresponds to a linear regression. Second line: error distribution histograms for (d) MEM, (e) S-SDM$_{binroc}$, and (f) S-SDM$_{resamp}$ for the Diablerets area.

Comparison between the two approaches

Despite the significantly different intercepts and slopes of their regression lines with the observed SR, the predictions from MEM and the three S-SDM approaches were strongly correlated with one another with values ranging between 0.84 for the correlation between S-SDM$_{binroc}$ and MEM to 0.99 for the correlation between S-SDM$_{sumprob}$ and S-SDM$_{resamp}$ (Table 1). Despite such high correlation, the S-SDM$_{binroc}$ tended to predict greater SR than did the direct MEM and the S-SDM$_{resamp}$ and S-SDM$_{sumprob}$.

Table 1: Spearman correlations between the various approaches to model species richness. S-SDM: stacked species distribution model, MEM: macroecological model. Predictions of S-SDMs were assembled in three ways: *binroc*: binarized predictions using a probability threshold; *sumprob*: sum of predicted probabilities; *resamp*: resampling trials from binomial distributions based on the predicted probabilities. See main text for explanations.

	Observed	MEM	S-SDM$_{binroc}$	S-SDM$_{sumprob}$	S-SDM$_{resamp}$
Observed	1				
MEM	0.674	1			
S-SDM$_{binroc}$	0.688	0.842	1		
S-SDM$_{sumprob}$	0.787	0.898	0.927	1	
S-SDM$_{resamp}$	0.787	0.898	0.926	0.998	1

Along the elevation gradient, the observed SR showed a hump-shaped curve with a peak of richness at mid-elevation, a decrease in SR toward high elevations and, to a lesser extent, a decrease in SR toward low elevations (Figure 2a). The MEM reproduced the hump-shaped curve well (Figure 2c). In contrast, the S-SDM$_{binroc}$ was not as successful in reproducing the hump-back curve, which was less pronounced (Figure 2b). Both the S-SDM$_{resamp}$ (Figure 2d) and the S-SDM$_{sumprob}$ (Figure S4c) were unable to reflect appropriately the hump-shaped curve. Note that some levels of overprediction were also observed with the S-SDM$_{resamp}$ and S-SDM$_{sumprob}$ at low elevations.

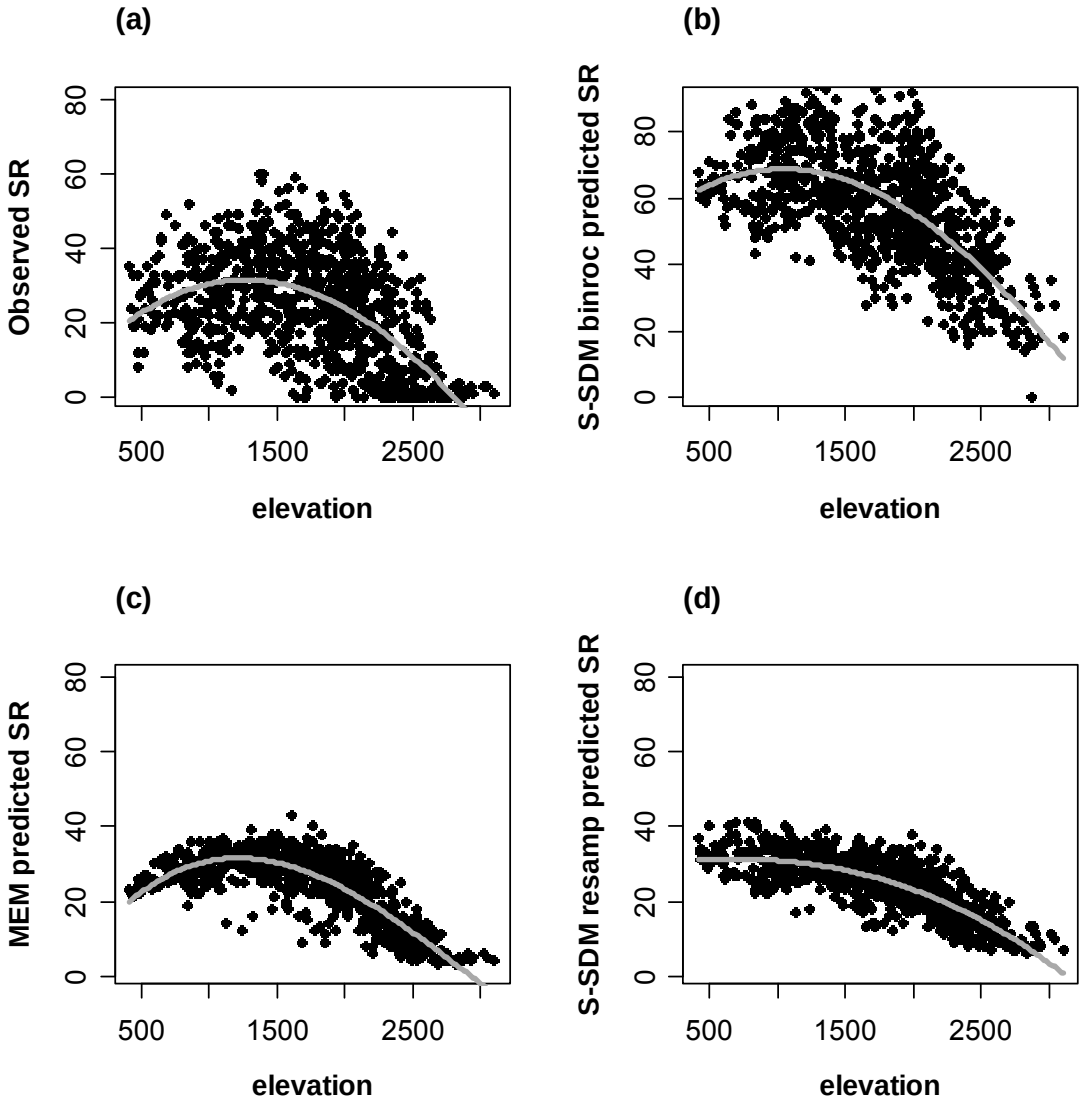


Figure 2: SR along the elevation gradient: (a) observed, (b) predicted by S-SDM$_{binroc}$, (c) predicted by MEM and (d) predicted by S-SDM$_{resamp}$ for the Diablerets area. The grey curve corresponds to a linear regression with quadratic terms between SR and elevation.

DISCUSSION

The aim of this study was to compare the two general approaches frequently used to model and predict species richness (SR): macroecological models (MEM) and stacked species distribution models (S-SDM), with three different S-SDM implementations. Because MEM and S-SDM rely on different theoretical assumptions, we expected discrepancies in their fitted and predicted SR patterns. We have shown that, depending on their implementation, these two modelling approaches yield different predictions, leading in most cases to the rejection of our null hypothesis of similar SR predictions. Although both methods are commonly used in the literature, the choice of one over the other is rarely addressed. Here, we discuss the strengths and weaknesses of MEM and the three different S-SDM options in light of our results.

Macroecological models

As expected, the MEM-predicted values were overall centred on the observed mean with errors symmetrically distributed around it. Hence, no bias was observed overall suggesting that this approach correctly captures the main spatial and environmental trends in SR. However, while this approach was accurate on average, it failed to reproduce the variability observed between plots within similar environmental conditions. Consequently, this approach was inappropriate for predicting extreme values, as shown in other MEM studies (e.g. Guisan & Theurillat, 2000a; Thuiller *et al.*, 2006; Algar *et al.*, 2009). This result can explain why the correlation between the modelled and observed richness was lower for MEM than for S-SDM. Algar *et al.* (2009) and Newbold *et al.* (2009) also reported the same trend (although in Algar, only for the MEM without a spatial autocorrelation term, as no spatial autocorrelation was included in their S-SDMs). Most importantly, MEM loses the information on species identities (early on, when the data are prepared).

Stacked species distribution models

The correlation between observed and predicted SR was higher for all S-SDMs than for MEM. The S-SDMs were more successful at scoring the plots according to their SR: plots with extremely high or low SR were predicted correctly relative to others. However, S-SDMs suffer from several drawbacks, depending on how mapped predictions are stacked. S-SDM$_{binroc}$ was less able to accurately predict absolute values of SR. Instead, the predictions were always biased toward SR overprediction. This result is consistent with other studies using S-SDMs with binarized maps, for

different taxonomic groups. Pineda & Lobo (2009) reported overestimation when predicting amphibian SR by S-SDM. Algar *et al.* (2009) showed that S-SDM tended to overpredict butterfly SR in species-rich areas. Finally, results from Newbold *et al.* (2009) showed that S-SDM for Egyptian butterflies and mammals overpredicted SR more than MEM for the same dataset. On the contrary, Lehmann *et al.* (2002) showed no clear differences between SR derived by summing raw probabilities predicted by SDMs and direct prediction of SR. This could indicate that the overprediction issue is mainly a consequence of the threshold used while building SR maps. Yet, even if the overprediction issue is drastically reduced with the $S\text{-}SDM_{resamp}$ and $S\text{-}SDM_{sumprob}$ for our dataset, these approaches tend to overpredict the richness in species-poor plots, an issue that still requires further investigation.

The results of $S\text{-}SDM_{resamp}$ and $S\text{-}SDM_{sumprob}$ are close to those obtained by MEM, with an even better correlation with the observed data. $S\text{-}SDM_{resamp}$ also shows an interesting ability to provide an estimate of prediction uncertainty. However, while these approaches therefore appear to offer a good way to predict unbiased SR values, they suffer from a practical drawback when one wishes to retrieve other properties of assemblages than SR, such as assemblage composition. A way forward could be to consider the assemblage trials from the binomial resampling approach as different possible end-points of assembly, as early suggested by Law & Morton (1993). However, are they all equiprobable? They might be statistically, but not from the perspective of ecological theory (e.g., assembly rules). If not, how do we assess their individual probability of realization? Are there some communities that are more constantly predicted among all trials? These questions require further investigation. Furthermore, in the case of the non-equiprobability of these end-points of assembly, rules must be added to select the most probable final assemblage. In summary, from the pure SR prediction perspective, the $S\text{-}SDM_{resamp}$ and $S\text{-}SDM_{sumprob}$ provide potentially powerful predictions, but with the same drawback as the MEM in the sense that they ultimately lose the list of co-occurring species.

Reproducing species richness patterns along the elevation gradient

To better understand the divergence between predictions made by the modelling approaches, we further compared them to the observed SR pattern along the elevation gradient. The observed SR shows a hump-shaped curve, with a peak of SR between 1,300 and 2,000 m. This pattern is consistent with those reported in several other studies (Minchin, 1989; Grytnes, 2003; Rahbek, 2005; Maurer *et al.*, 2006). This peak of SR could result from several distinct phenomena, such as the mid domain

effect (Colwell & Lees, 2000), where species ranges thrown randomly within a bounded range are expected to overlap more in the centre of the domain than at the borders. The intermediate disturbance hypothesis (Connell, 1978) also suggests that the maximum SR in our study area should occur at middle elevations due to less optimal conditions occurring at both ends of the gradient, with intensive land use and fertilization at low elevations limiting habitat types and strengthening plant competition, leading to species exclusion (Eriksson *et al.*, 1995; Foster & Gross, 1998) and to severe environments at high elevations limiting the number of adapted species. A third, alternative, explanation could be the overlap of subalpine and alpine species pools, causing an increase in SR at the ecotone between these two vegetation belts (Grytnes, 2003).

The MEM was better able to reproduce the hump-shaped curve seen in observed data. Other factors not taken into account in the model, such as human influences that are particularly important at lower elevations (e.g., intensive land use) or natural disturbance regimes that are more intense at higher elevations (Randin *et al.*, 2009a), were probably indirectly accounted for because they are correlated with temperature. The overall pattern along the elevation gradient was less apparent with the three S-SDMs. S-SDM$_{resamp}$ and S-SDM$_{sumprob}$ performed particularly badly in this respect, as they missed the mid-elevation optimum and predicted a rather linear decrease in SR from the lowest to highest elevations, which was likely caused by the tendency to overpredict SR at lower elevations.

Conclusion and perspectives

We have shown that MEM and S-SDMs have different strengths and weaknesses. MEM predicts a realistic number of species on average and better reproduces the observed hump-shaped SR pattern along the elevation gradient, but its predictions are overall less correlated with observed SR. This approach seems therefore more appropriate for testing biogeographic hypotheses and studying general patterns of SR, because it avoids overprediction, but it does not provide any information on the identity of species and community composition.

The results are more conflicting among the three types of S-SDMs. S-SDM$_{binroc}$, despite its strong overprediction bias, is better correlated overall with the observed SR than MEM (but less so than the other two S-SDM approaches; see below). Moreover, it presents a significant practical advantage over the two other S-SDMs and MEM in providing a final prediction of species composition. However, we showed that, despite

a good correlation with observations, S-SDM$_{binroc}$ largely overpredicts SR, and a solution must be found to limit the predicted SR.

S-SDM$_{sumprob}$ and S-SDM$_{resamp}$ both show very good correlations with observed data. Although resulting from the same theoretical approach as S-SDM$_{binroc}$, they produce predictions that share the similarity with MEM of being unbiased. However, in contrast to MEM they reflect imperfectly the altitudinal pattern and in contrast to S-SDM$_{binroc}$ they do not provide a single unequivocal final species composition.

For greater perspective, it would be interesting to assess in more detail the multiple assemblage compositions predicted by the binomial resampling trials, especially to investigate whether a processing of these could allow a single final prediction or if the competing alternative end points of assembly must be considered as the final prediction itself. It would also be interesting to assess more thoroughly how the larger species pool yielded by the roc-binarized approach relates to the multiple end points predicted by the resampling approach (e.g., are the most frequent assemblages predicted by the latter all subsets of the larger roc-binarized species pool?).

Therefore, despite its overprediction bias, S-SDM$_{binroc}$ appears to be the most appropriate method when a single species list is needed, provided a solution can be implemented for further selecting the species entering the realized community. How can this be done? As proposed by Guisan A. & Rahbek C. (in press; see also Ferrier & Guisan 2006, figure S1 in supplementary material), MEM and S-SDM$_{binroc}$ could be combined to take advantage of their respective strengths. Without constraints, the most commonly used S-SDM$_{binroc}$ predicts the list of all species potentially able to occur under given environmental conditions. The species number predicted by traditional MEM corresponds to the mean number of species that can actually occur at a place given macroecological conditions, but the maximum richness could also be predicted through quantile regression (Cade & Noon, 2003). The final list of species predicted for a location could be obtained by constraining the S-SDM$_{binroc}$ prediction with a maximum SR value predicted by a separate MEM, and the restricted number of species allowed to coexist may be selected through a set of ecological assembly rules, such as those based on checkerboard and co-occurrence patterns (Gotelli, 2000) or functional traits (Wilson & Roxburgh, 1994).

ACKNOWLEDGEMENTS

We thank Simon Ferrier and three anonymous referees for very useful comments on a previous version of the manuscript and everyone who helped with data collection. This study was supported by the Swiss National Science Foundation (grant Nr. 31003A-125145 to A. Guisan) and by the European Commission (ECOCHANGE project). The Belalp data were obtained within the framework of the Priority Programme Environment funded by SNF (grant Nr. 5001-03504).

REFERENCES

Aarssen, L.W. & Schamp, B.S. (2002) Predicting distributions of species richness and species size in regional floras: Applying the species pool hypothesis to the habitat templet model. *Perspectives in Plant Ecology Evolution and Systematics*, 5, 3-12.

Algar, A.C., Kharouba, H.M., Young, E.R. & Kerr, J.T. (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography*, 32, 22-33.

Bouët, M. (1985) *Climat et météorologie de la Suisse romande*. Payot, Lausanne.

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

Cade, B.S. & Noon, B.R. (2003) A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1, 412-420.

Colwell, R.K. & Lees, D.C. (2000) The mid-domain effect: geometric constraints on the geography of species richness. *Trends in Ecology & Evolution*, 15, 70-76.

Connell, J.H. (1978) Diversity in tropical rain forests and coral reefs - high diversity of trees and corals in maintained only in a non-equilibrium state. *Science*, 199, 1302-1310.

Cornell, H.V. & Lawton, J.H. (1992) Species interactions, local and regional processes, and limits to the richness of ecological communities - A theoretical perspective. *Journal of Animal Ecology*, 61, 1-12.

Cumming, G.S. (2000) Using habitat models to map diversity: pan-African species richness of ticks (Acari : Ixodida). *Journal of Biogeography*, 27, 425-440.

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.

Dufour, A., Gadallah, F., Wagner, H.H., Guisan, A. & Buttler, A. (2006) Plant species richness and environmental heterogeneity in a mountain landscape: effects of variability and spatial configuration. *Ecography*, 29, 573-584.

Engler, R., Randin, C.F., Vittoz, P., Czaka, T., Beniston, M., Zimmermann, N.E. & Guisan, A. (2009) Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? *Ecography*, 32, 34-45.

Eriksson, A., Eriksson, O. & Berglund, H. (1995) Species abundance patterns of plants in swedish seminatural pastures. *Ecography*, 18, 310-317.

Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, 43, 393-404.

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.

Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, 12, 42-58.

Foster, B.L. & Gross, K.L. (1998) Species richness in a successional grassland: Effects of nitrogen enrichment and plant litter. *Ecology*, 79, 2593-2602.

Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29, 1189-1232.

Friedman, J.H., Hastie, T. & Tibshirani, R. (2000) Additive Logistic Regression: a Statistical View of Boosting. *Annals of Statistics*, 28, 337-374.

Gelfand, A.E., Schmidt, A.M., Wu, S., Silander, J.A., Latimer, A. & Rebelo, A.G. (2005) Modelling species diversity through species level hierarchical modelling. *Journal of the Royal Statistical Society Series C-Applied Statistics*, 54, 1-20.

Gioia, P. & Pigott, J.P. (2000) Biodiversity assessment: a case study in predicting richness from the potential distributions of plant species in the forests of south-western Australia. *Journal of Biogeography*, 27, 1065-1078.

Gleason, H. (1926) The Individualistic Concept of the Plant Association. *Bulletin of the Torrey Botanical Club*, 53, 7-26.

Gotelli, N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology*, 81, 2606-2621.

Gould, W. (2000) Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological Applications*, 10, 1861-1870.

Grytnes, J.A. (2003) Species-richness patterns of vascular plants along seven altitudinal transects in Norway. *Ecography*, 26, 291-300.

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

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

Guisan, A. & Theurillat, J.-P. (2000b) 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. & Rahbek, C. (in press) SESAM - A new framework for predicting spatio-temporal patterns of species assemblages: Integrating macroecological and species distribution models. *Journal of Biogeography*,

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.

Hastie, T. & Tibshirani, R. (1990) *Generalized Additive Models*. Chapman and Hall, London.

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guegan, J.F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E. & Turner, J.R.G. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84, 3105-3117.

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

Huston, M. (1979) General hypothesis of species-diversity. *American Naturalist*, 113, 81-101.

Hutchinson, G.E. (1957) Population studies - animal ecology and demography - concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415-427.

Jetz, W. & Rahbek, C. (2002) Geographic range size and determinants of avian species richness. *Science*, 297, 1548-1551.

Körner, C. (1999) *Alpine plant life*. Springer, Heidelberg.

Law, R. & Morton, R.D. (1993) Alternative permanent states of ecological communities. *Ecology*, 74, 1347-1361.

Lehmann, A., Leathwick, J.R. & Overton, J.M. (2002) Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodiversity and Conservation*, 11, 2217-2238.

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

Loreau, M. (2000) Are communities saturated? On the relationship between alpha, beta and gamma diversity. *Ecology Letters*, 3, 73-76.

Luoto, M., Virkkala, R., Heikkinen, R.K. & Rainio, K. (2004) Predicting bird species richness using remote sensing in boreal agricultural-forest mosaics. *Ecological Applications*, 14, 1946-1962.

Mackey, R.L. & Currie, D.J. (2000) A re-examination of the expected effects of disturbance on diversity. *Oikos*, 88, 483-493.

Mackey, R.L. & Currie, D.J. (2001) The diversity-disturbance relationship: Is it generally strong and peaked? *Ecology*, 82, 3479-3492.

Maurer, K., Weyand, A., Fischer, M. & Stocklin, J. (2006) Old cultural traditions, in addition to land use and topography, are shaping plant diversity of grasslands in the Alps. *Biological Conservation*, 130, 438-446.

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

Michalet, R., Brooker, R.W., Cavieres, L.A., Kikvidze, Z., Lortie, C.J., Pugnaire, F.I., Valiente-Banuet, A. & Callaway, R.M. (2006) Do biotic interactions shape both sides of the humped-back model of species richness in plant communities? *Ecology Letters*, 9, 767-773.

Minchin, P.R. (1989) Montane vegetation of the Mt Field Massif, Tasmania - A test of some hypotheses about properties of community patterns. *Vegetatio*, 83, 97-110.

Moser, D., Dullinger, S., Englisch, T., Niklfeld, H., Plutzar, C., Sauberer, N., Zechmeister, H.G. & Grabherr, G. (2005) Environmental determinants of vascular plant species richness in the Austrian Alps. *Journal of Biogeography*, 32, 1117-1127.

Newbold, T., Gilbert, F., Zalat, S., El-Gabbas, A. & Reader, T. (2009) Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography*, 36, 2085-2095.

Nogues-Bravo, D., Araujo, M.B., Romdal, T. & Rahbek, C. (2008) Scale effects and human impact on the elevational species richness gradients. *Nature*, 453, 216-U8.

Palmer, M.W. & Dixon, P.M. (1990) Small-scale environmental heterogeneity and the analysis of species distributions along gradients. *Journal of Vegetation Science*, 1, 57-65.

Parviainen, M., Marmion, M., Luoto, M., Thuiller, W. & Heikkinen, R.K. (2009) Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. *Biological Conservation*, 142, 2501-2509.

Pausas, J.G. & Austin, M.P. (2001) Patterns of plant species richness in relation to different environments: An appraisal. *Journal of Vegetation Science*, 12, 153-166.

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

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

Rahbek, C. (2005) The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, 8, 224-239.

Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 4534-4539.

Randin, C.F., Vuissoz, G., Liston, G.E., Vittoz, P. & Guisan, A. (2009a) Introduction of Snow and Geomorphic Disturbance Variables into Predictive Models of Alpine Plant Distribution in the Western Swiss Alps. *Arctic Antarctic and Alpine Research*, 41, 347-361.

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.

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

Ricklefs, R.E. (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, 7, 1-15.

Ricklefs, R.E. (2008) Disintegration of the Ecological Community. *American Naturalist*, 172, 741-750.

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

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. & Wall, D.H. (2000) Biodiversity - Global biodiversity scenarios for the year 2100. *Science*, 287, 1770-1774.

Soberon, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10, 1115-1123.

Steck, A. (1983) Geologie der Aletschregion. *Bulletin de la Murithienne*, 101, 135-154.

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.

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

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

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.

Vincent, P.J. & Haworth, J.M. (1983) Poisson regression models of species abundance. *Journal of Biogeography*, 10, 153-160.

Whittaker, R.H. (1972) Evolution and Measurement of Species Diversity. *Taxon*, 21, 231-251.

Whittaker, R.J., Nogues-Bravo, D. & Araujo, M.B. (2007) Geographical gradients of species richness: a test of the water-energy conjecture of Hawkins et al. (2003) using European data for five taxa. *Global Ecology and Biogeography*, 16, 76-89.

Wilson, J.B. & Roxburgh, S.H. (1994) A Demonstration of guild-based assembly rules for a plant community, and determination of intrinsic guilds. *Oikos*, 69, 267-276.

SUPPLEMENTARY INFORMATIONS

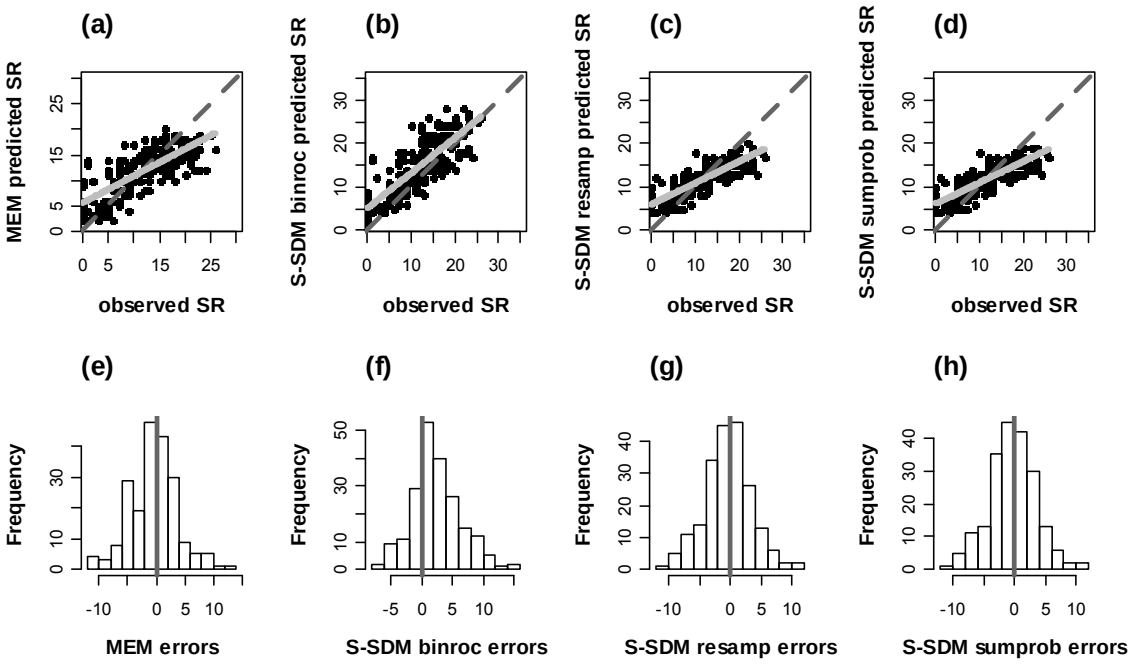


Figure S1: Results for the Belalp study area. First line: predicted species richness plotted against the observed data for (a) MEM, (b) S-SDM$_{binroc}$, (c) S-SDM$_{resamp}$, (d) S-SDM$_{sumprob}$. The dashed line shows a perfect correspondence between observed and predicted SR and the plain line corresponds to a linear regression. Second line: error distribution histograms for (e) MEM, (f) S-SDM$_{binroc}$, (g) S-SDM$_{resamp}$, (h) S-SDM$_{sumprob}$.

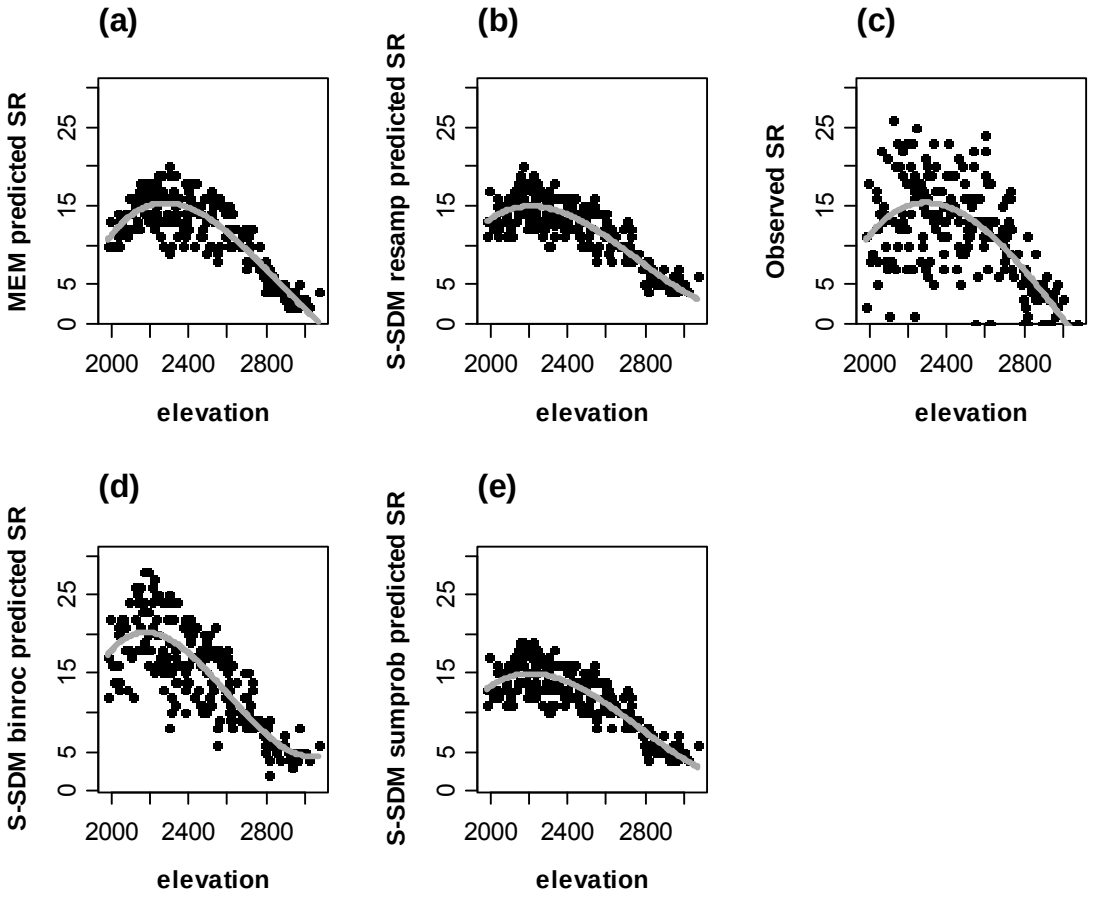


Figure S2: Results for the Belalp study area. Species richness (counts) along the elevation gradient: (a) predicted by MEM, (d) predicted by S-SDM$_{binroc}$, (b) predicted by S-SDM$_{resamp}$, (e) predicted by S-SDM$_{sumprob}$ and (c) observed.

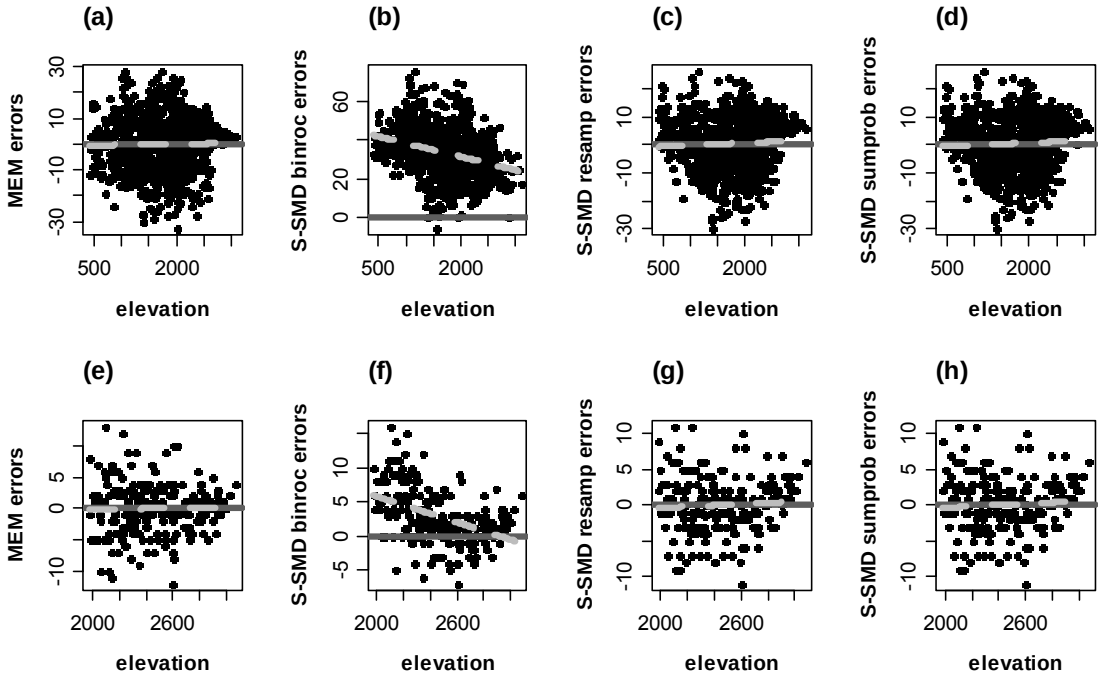


Figure S3: Errors (predicted – observed values) of: (a) MEM, (b) S-SDM$_{binroc}$, (c) S-SDM$_{resamp}$, and (d) S-SDM$_{sumprob}$ plotted along elevation for the Diablerets area; (e), (f), (g) and (h) display the same for Belalp area. The dashed line is a trend line.

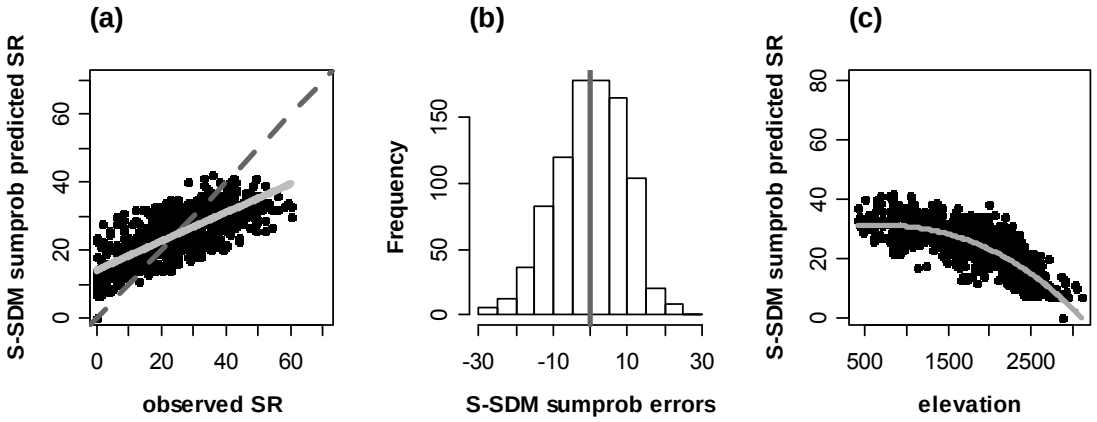


Figure S4: Plots showing the results of the SR obtained by summing the probabilities for the Diablerets area. (a) predicted SR plotted against the observed data, (b) error distribution histograms and (c) SR along the elevation gradient.

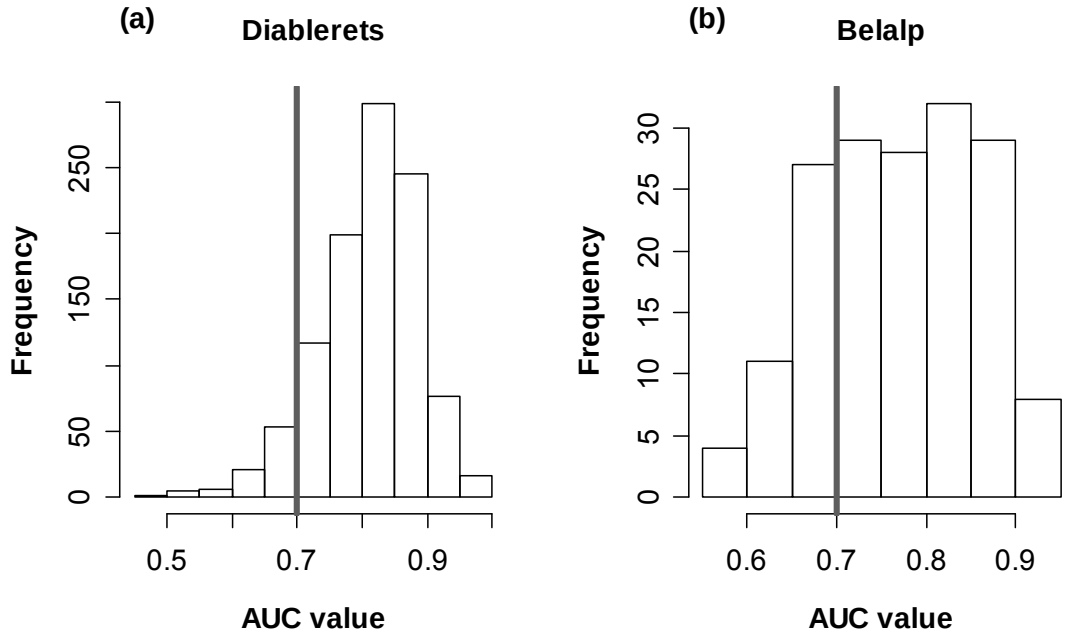


Figure S5: Histogram of AUC values for (a) Diableret and (b) Belalp areas. Models with an AUC higher than 0.7 can be considered as useful.

CHAPTER 4

Predicting current and future community patterns of plant functional traits

Anne Dubuis [1]
Leila Rossier [1]
Julien Pottier [2]
Loïc Pellissier [1]
Pascal Vittoz [1]
Antoine Guisan [1,3]

1 Department of Ecology and Evolution, University of Lausanne, Biophore building, 1015 Lausanne, Switzerland

2 INRA, Grassland Ecosystem Research Unit (UREP), 5 Chemin de Beaulieu, 63100 Clermont-Ferrand, France

3 Faculty of Geoscience and the Environment, University of Lausanne, 1015 Lausanne, Switzerland

This paper is in press in Ecography
DOI: 10.1111/j.1600-0587.2013.00237.x

Contribution to the project: I participated to the field and lab work for plant, functional traits and soil data collection, I performed the statistical analysis. I wrote the initial draft of the paper and lead the reviewing process.

ABSTRACT

Community-level patterns of functional traits relate to community assembly and ecosystem functioning. By modelling the changes of different indices describing such patterns - trait means, extremes and diversity in communities - as a function of abiotic gradients, we could understand their drivers and build projections of the impact of global change on the functional components of biodiversity. We used five plant functional traits (vegetative height, specific leaf area, leaf dry matter content, leaf nitrogen content and seed mass) and non-woody vegetation plots to model several indices depicting community-level patterns of functional traits from a set of abiotic environmental variables (topographic, climatic and edaphic) over contrasting environmental conditions in a mountainous landscape. We performed a variation partitioning analysis to assess the relative importance of these variables for predicting patterns of functional traits in communities, and projected the best models under several climate change scenarios to examine future potential changes in vegetation functional properties. Not all indices of trait patterns within communities could be modelled with the same level of accuracy: the models for mean and extreme values of functional traits provided substantially better predictive accuracy than the models calibrated for diversity indices. Topographic and climatic factors were more important predictors of functional trait patterns within communities than edaphic predictors. Overall, model projections forecast an increase in mean vegetation height and in mean specific leaf area following climate warming. This trend was important at mid elevation particularly between 1000 and 2000 m asl. With this study we showed that topographic, climatic and edaphic variables can successfully model descriptors of community-level patterns of plant functional traits such as mean and extreme trait values. However, which factors determine the diversity of functional traits in plant communities remains unclear and requires more investigations.

Key-words: climate change, community assembly, community weighted mean, functional diversity, functional traits, modelling, Swiss Alps, vegetation.

INTRODUCTION

Functional traits provides better generality in understanding and predicting the formation and structure of plant communities as well as ecosystem functions than approaches based on species identity alone (Keddy 1992b, Keddy 1992a, Diaz and Cabido 2001, Hooper et al. 2005, McGill et al. 2006). Functional traits are defined as any morphological, physiological or phenological features measurable at the individual level that affect individual performances (Violle et al. 2007). By contrasting trait values for individual species to the ones aggregated at the community level (Diaz *et al.* 2007), functional traits are supposed to enable the refinement of predictions of communities composition along environmental gradients (Shipley et al. 2006, Douma et al. 2012), but also to bring more understanding in the modification of ecosystem functioning (Lavorel and Garnier 2002, de Bello et al. 2010, Van Bodegom et al. 2012). Refined predictions of communities in space as well as a better insight in the evolution of ecosystem functioning are highly desirable in the current context of global changes (Nogues-Bravo and Rahbek 2011). Yet, few studies have investigated whether measures of aggregated traits at the community level can be modelled and predicted in space (Pellissier et al. 2010b, Sonnier et al. 2010b).

There are three main expressions of the distribution of trait values at the community level, which consider the mean (or median), the extremes in trait values or their diversity. First, the mean of the trait values weighted by the respective abundance of each species (community weighted mean ; CWM; Garnier et al. 2004) has been used to functionally characterize plant communities in different environments (Cornwell and Ackerly 2009, Venn et al. 2011) so as to ultimately better understand community assembly (Ackerly and Cornwell 2007). When used to study ecosystem functions, CWM reflects the mass ratio hypothesis (Grime 1998), which proposes that the dominant trait value in the community has the greatest impact on ecosystem functioning. Second, extremes in trait values are simply measured by taking either minimum/maximum or some quantiles (e.g. 5^{th} and 95^{th}) of the distribution of trait values in the community. These values may reflect interesting filtering effects (Keddy 1992), showing which limiting trait value allows a species to be included in a community in a given environment. To our knowledge, extreme values of functional traits have not yet been investigated in this context. Thirdly, the diversity of trait values (or functional diversity) quantifies the breadth of occupied trait space by species inside a community, and the distribution of species or individuals within that trait space (Schleuter et al. 2010). It can be estimated using several indices, such as richness,

divergence or evenness (Mouchet et al. 2010). Functional diversity metrics allow the identification of convergence-divergence assembly rules in plant communities (Cornwell and Ackerly 2009, de Bello et al. 2009) or to assess ecosystem functioning, through the functional complementarity hypothesis (Hector et al. 1999, Symstad 2000, Diaz and Cabido 2001). The latter states that coexisting species with different functional attributes exhibit complementary resource use (Tilman 1997). Therefore, functional properties of plant communities are useful to characterize biodiversity, community structure and ecosystem functioning.

Unfortunately, this information is only available for communities in which the total vegetation is known and where trait data are available for most of the species. Such data are rarely available in a spatially-explicit way across large geographic areas and large environmental gradients, making it difficult to assess current and future spatial patterns of community functional properties over large areas (but see Swenson and Weiser 2010, Swenson et al. 2012). Several reviews have emphasised the need to better study the relation between functional traits and environmental gradients (Weiher et al., 2011) and to generalise functional data to an entire geographic region (Moles et al. 2008, Webb et al. 2010, Van Bodegom et al. 2012). It remains therefore crucial to understand community functional patterns along environmental gradients and test if and which functional community properties can be predicted in space and in time, e.g. according to global change scenarios. Previous studies have already used empirical data and statistical models to predict the relative abundance of discrete traits in space (Pellissier et al. 2010b) and time (Küster et al. 2010), and to predict the trait composition of communities from stress and disturbance gradients (Sonnier et al. 2010a). However, to our knowledge, no attempt has been made to (i) assess whether means, extremes and diversity of communities' functional traits respond to environmental variables, and (ii) use the modelled relationships to predict the distribution of these trait metrics in space as a function of the selected environmental predictors.

In this paper we aim at addressing this challenge. We used vegetation data sampled across a wide altitudinal gradient to relate descriptors of community-level patterns of plant functional traits (several indices summarizing traits means, trait range and trait diversity) to environmental variables with statistical models. We focused our analyses on key functional traits associated with plant performance in the growth, persistence and reproduction phase of their life cycle according to the leaf-height-seed plant strategy (LHS) of Westoby (Westoby, 1998). We used three leaf traits (specific leaf

area, leaf dry matter content and leaf nitrogen content), plant height and seed mass. We first related the community functional indices individually to the environmental gradients of our study area. Second, we built models of functional indices as a function of environmental predictors and evaluated their ability to predict functional indices in space. Then, to disentangle the respective importance of edaphic and topo-climatic predictors in the models of functional properties, we applied a variation partitioning analysis. Finally, to obtain insights of the fate of the vegetation of our study area, we projected our models in the future considering two climate change scenarios for two periods ending in 2040 and 2070.

MATERIALS AND METHODS

Study area

The study area is located in the Western Swiss Alps (Canton de Vaud, Switzerland, 46°10' to 46°30'N, 6°50' to 7°10'E) and covers 700 km^2, with an elevation ranging from 375 to 3210 m. The general climate of this mountain region is temperate. 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. Human impacts occur frequently in this region up to the subalpine area and are characterised by agricultural activities, including the addition of fertiliser, mowing and grazing (see Randin et al. 2009 for more details).

Vegetation data

The species occurrence data from the vegetation plots used in our analysis originate from fieldwork conducted between 2002 and 2009 in the study area (Randin et al. 2009, Dubuis et al. 2011). In total, 912 vegetation plots of 4 m^2 each were inventoried according to a random-stratified sampling design based on plots elevation, slope and aspect. This sampling was limited to open, non-woody vegetation (grassland, meadow, rock and scree). All of the vascular plant species were inventoried with their relative abundance recorded according to an ordinal scale (Vittoz and Guisan 2007), where 0: absent; 1: ≤ 1%; 2: 1-5%; 3: 5-12.5 %; 4: 12.5-25 %; 5: 25-50%; 6: 50-75%; and 7: >75 %. We used the median of these classes for further calculations.

For further analyses, we retained only the 605 plots for which trait data were available for more than 80% of the relative vegetation cover (Pakeman and Quested 2007). Among these plots, 195 had edaphic data available (see below for details on the edaphic variables). The subset of 195 plots covers the same geographic extent as the full set of plots and is also fully random-stratified with regard to elevation, slope and orientation.

Trait data collection

The 240 most abundant angiosperm species in the above dataset were sampled. Between four and 20 individuals per species were measured in sites with contrasting environmental conditions so as to cover as much of the total distribution range of each species as possible. We followed the growing season according to altitude and sampled all of the individuals of a species at the same phenological stage whenever possible. One well-developed entire leaf was collected per individual (Cornelissen et al.

2003) for trait measurement. We used the average trait value among all sampled individuals for each species for further analyses.

Vegetative height (H) was measured in the field (in mm) as the distance between the top of the photosynthetic tissue and the ground. This trait is related to competitive ability and is correlated with above-ground biomass (Cornelissen et al. 2003). Specific leaf area (SLA) was calculated as the ratio of the leaf surface to its dry mass (without the leaf petiole) and expressed in mm^2*mg^{-1}. SLA is correlated with the relative growth rate and photosynthetic ability of herbaceous species (Cornelissen et al. 2003). Leaf dry matter content (LDMC) is the ratio of leaf dry mass to its saturated fresh mass (in $mg\ g^{-1}$) and was measured using the partial-rehydration method described by Varietti *et al.* (2007). LDMC is related to the average density of the leaf tissues and is negatively correlated with the potential relative growth rate, but it is positively correlated with leaf lifespan (Cornelissen et al. 2003) and resource-use strategy (Wilson et al. 1999). Leaf nitrogen content (LNC) was analysed by combustion, using an elemental analyser on one sample of mixed ground leaves per species. LNC is the total amount of nitrogen per unit of leaf dry mass and is linked to the photosynthetic rate of a plant (Cornelissen et al. 2003). Seed mass (SM, in mg) data originate from literature and field measurements (for origin of data, see Pellissier et al. 2010a) and is expressed in milligrams. Seed mass represents trade-offs in the plant investment in reproduction. Smaller seeds can be produced in higher number but contain limited resources, while heavier, less numerous seeds contain more resources useful for the seedling to survive. These five traits stand for the three functional axes described in Westoby's (1998) leaf-height-seed plant strategy.

Functional indices

Several indices were calculated from the five traits (H, SLA, LDMC, LNC, SM), and these indices were used to depict trait patterns at the community level. First, we computed the community mean of the trait weighted by the abundance of each species (CWM, Garnier et al. 2004). We computed the 5^{th} and 95^{th} quantiles of trait values in each plot as indication of trait extreme values. We use these values to reflect minimal and maximal trait values without being distorted by potential outliers. Finally, regarding diversity indices, we computed: functional richness (FRic) representing the amount of functional trait space filled by the community, functional evenness (FEve) corresponding to the evenness of the abundance distributions in the functional trait space, and Rao's quadratic entropy (de Bello et al. 2009), calculated as the mean functional distance between species when weighted by their relative abundance. The

use of these indices is recommended in a comparative study conducted by Mouchet et al. (2010).

These six indices were each computed for the five traits. As recommended for situations of non-normality, all traits were log transformed before the functional indices were calculated. We computed the correlations between all indices and species richness to assess their relationships. CWM, FRic and FEve were computed with the FD package (Laliberte and Legendre 2010). Rao was computed with the R codes provided in de Bello et al. (2009) in the R software environment (R Development Core Team 2011).

Environmental predictors

We used three groups of environmental variables for our analysis. The first group was composed of topographic and climatic variables (hereafter called topo-climatic predictors); it included degree-days, moisture index, global solar radiation, slope and topographic position. These data were generated at a resolution of 25 m from a digital elevation model and temperature and precipitation data (the average between 1961 and 1990) recorded by the Swiss network of meteorological stations. Degree-days are defined as the sum of days of the growing season (June, July and August) multiplied by the temperature above 0°C. The moisture index is defined as the difference between precipitation and potential evapotranspiration and represents the potential amount of water available in the soil at a site. We use its mean for the growing season. The potential global solar radiation was calculated over the year. The slope in degrees was derived from the digital elevation model with ArcGis 9.3 spatial analyst tool (ESRI 2008). The topographic position is an integration of topographic features at various scales and is computed with moving windows. Positive values of this variable indicate ridges and tops, whereas negative values correspond to valleys and sinks. The methods for computing these predictors are described in more detail in (Zimmermann and Kienast 1999).

Two groups of edaphic predictors comprised chemical variables (soil water pH, total nitrogen content, total phosphorus content) and physical variables (soil texture, *i.e.*, the percentages of clay and sand in the soil). These data were measured from soil samples taken from the first 10 cm of the organo-mineral horizon after removing the first organic horizon (see Dubuis et al. (In press) for more details on edaphic variables). The data were only recorded for the 195 most recent vegetation plots and were not available as continuous map; consequently, they cannot be used for model projections.

The correlations between all the predictors are presented in Supplementary material Appendix 1, Table A1.

Climate change scenarios

Future predictions for temperature and precipitation were made at a 100 m resolution and were averaged for 30 years periods ending in 2040 and 2070. The predictions were derived from the General Circulation Model of the Hadley Centre for Climate Prediction and Research (HadCM3). Two greenhouse gas emission scenarios were examined: A1FI and B2 (IPCC SRES, Nakicenovic and Swart 2000). A1FI is the most pessimistic scenario and predict a temperature increase of around 7°C over the study area, and B2 is a scenario with intermediate values and predict an increase of around 4°C for the study area. Future degree-days and moisture index data were calculated on the basis on the future temperature and precipitations maps.

Functional indices prediction

To understand the patterns of community functional indices along the chosen environmental predictors, we related each functional index to combinations of all environmental predictors (linear and quadratic) form with a linear regression model.

The spatial distribution of each index was then modelled with generalised additive models (GAM, Hastie and Tibshirani 1990) using a Gaussian distribution as implemented in the "gam" library in R (R Development Core Team 2011). Up to four degrees of freedom were allowed for the gam smoothing function. Using an exploratory approach, we ran four types of models for each index; models with topo-climatic predictors (TC), models with topo-climatic and chemical soil predictors (TC-CH), models with topo-climatic and physical soil predictors (TC-PS) and models with topo-climatic, chemical and physical soil predictors (TC-CH-PS). These analyses were applied to the 195 plots for which all predictors were available. The fit of the model was estimated by calculating the adjusted geometric mean-squared improvement – R^2 (Nagelkerke 1991) – which corresponds to the R^2 rescaled between 0 and 1 and adjusted for the number of observations and predictors in the model.

$$R^2 = 1 - \left(\left[n-1\right]/\left[n-p\right]\right) \times \left(1 - \left[1 - (L_0/L_M)^{2/n}\right]/\left[1 - (L_0)^{2/n}\right]\right)$$

Where, *n* is the total sample size, *p* is the total number of predictors used in the model, L_0 = the likelihood function for the model containing only the intercept and L_M is the likelihood function for the model containing all the predictors.

The predictive power of the models was evaluated with a repeated split-sample procedure using 100 repetitions. An evaluation dataset was obtained by randomly splitting the original dataset into two partitions of 70% and 30%, using the 70% partition for fitting the models and the remaining 30% partition for independently evaluating them. For each split-sample repetition and for each model, a Spearman rank correlation was calculated between the observed and predicted index value for the evaluation data set. The final evaluation value is the average of the 100 correlations.

Importance of environmental variables

To understand the relative importance of the three classes of predictors (topo-climatic variables and the chemical and physical attributes of the soil) in the TC-CH-PH models, we applied variation partitioning based on the partial correlation analyses. This approach partitions the variation into the following eight identifiable fractions of deviance: (1) pure topo-climatic (TC), (2) pure chemical attributes of the soil (CH), (3) pure physical attributes of the soil (PS), (4) shared chemical and physical properties of the soil (CH-PS), (5) shared topo-climatic and chemical properties of the soil (TC-CH), (6) shared topo-climatic and physical properties of the soil (TC-PS), (7) a combination of three classes (TC-CH-PS) and (8) unexplained variation (Borcard et al. 1992).

Projection under climate change conditions

Because continuous soil maps do not exist for our study area, the projections used a new set of GAMs calibrated only with topo-climatic predictors, thus assuming soil properties to remain constant. The 605 plots were used to calibrate these models at a 100 m resolution. Evaluation of the models was carried out with a repeated split sample procedure as explained above. Projections of the functional indices values were generated under current climate conditions and two climatic scenarios (A1FI and B2) for two time periods (2040 and 2070). The predicted proportion of change was calculated for each index and examined along the elevation gradient.

RESULTS

Patterns of community traits along environmental gradients

Community H and SLA increased with temperature and decreased with topographic position (Fig. 1 and Supplementary material Appendix 1, Table A2). Vegetation was overall smaller, with tougher leaves in cold conditions and on ridges. We found that the functional richness of SLA decreased as temperature increased, indicating increased convergence in warmer conditions (Fig. 3).

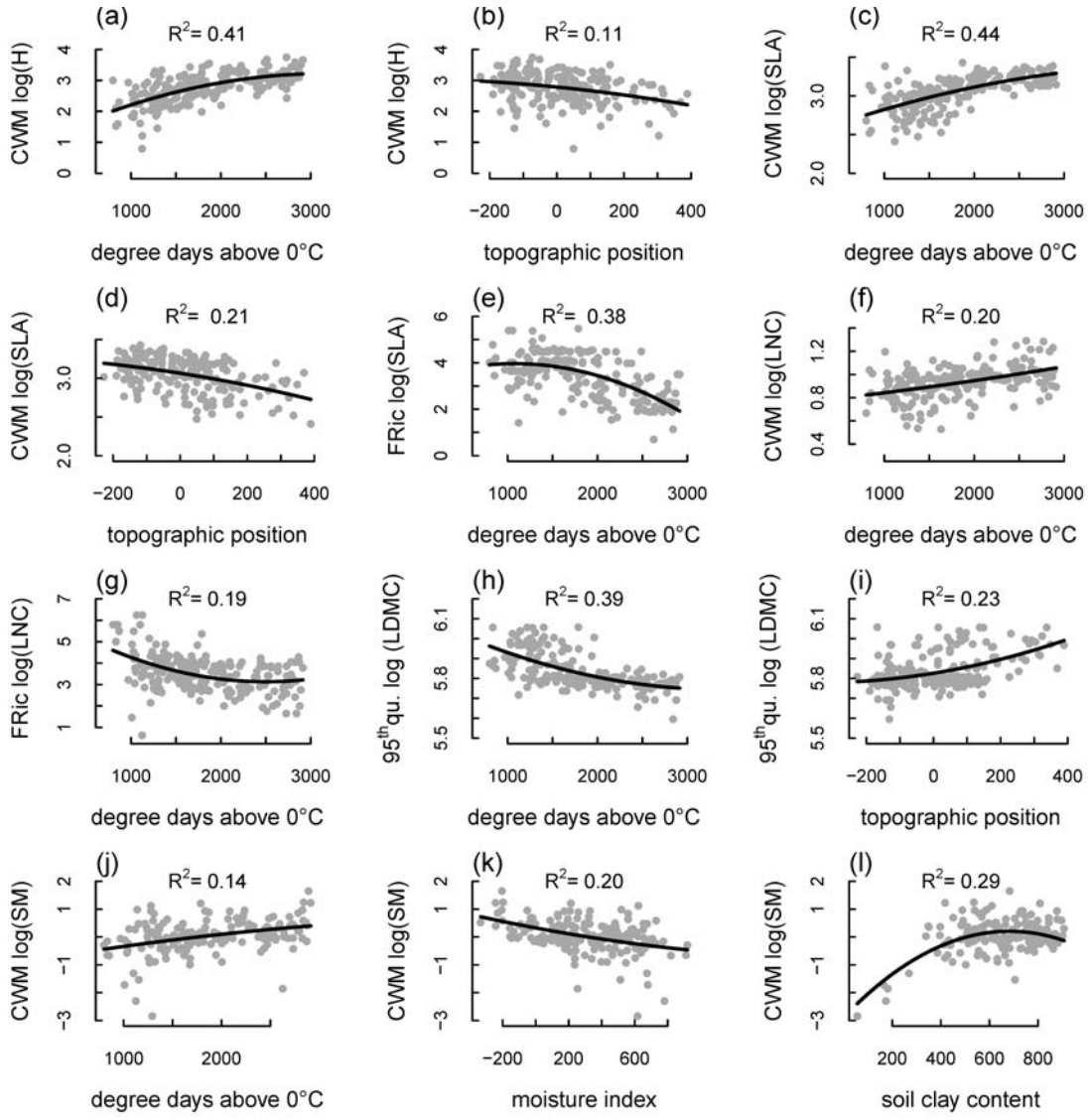


Figure 1. Scatter plots of the relationships between the functional indices and the environmental predictors that are significant and display the highest R^2. Lines indicate the fitted quadratic curve.

The community LNC increased with temperature while the LNC richness of communities decreased. The 95th quantile showed the strongest response to environmental gradient for LDMC: it decreased with temperature and increased with topographic position. Finally, community seed mass tended to be heavier in warmer and drier conditions and on soils with an high clay content (Fig. 1 and Supplementary material Appendix 1, Table A2).

Functional indices

The models better predicted functional trait means and extreme values (CWM, 5th and 95th quantile) than functional trait diversity (Table 1). Among the diversity indices, FRic showed the strongest correlation with the observed data; it had a maximum correlation coefficient of 0.66 between the predicted and observed values and a maximum R^2 of 0.61. FEve for all of the traits was poorly related with all combinations of the predictors (maximum correlation coefficients of 0.17, maximum R^2 of 0.31). Rao yielded intermediate results. Concerning the traits examined, the models for the SLA indices yielded the best results, (maximum correlation coefficients of 0.82 for 95^{th} quantile). H obtained a good fit and strong correlation for CWM, 5^{th} quantile and 95^{th} quantile, but the diversity indices for this trait obtained much lower values. LDMC, LNC and SM showed a lower fit and predictive accuracy for all indices (Table 1). The relationship between the functional indices and species richness are presented in Supplementary material Appendix 1, Table A3. FRic appear to be the index most correlated with species richness for all functional traits.

Table 1. R^2 and Spearman correlations between the observation and values predicted by the models with topo-climatic predictors (TC), topo-climatic and chemical edaphic predictors (TC-CH), topo-climatic and edaphic physical predictors (TC-PH) and the combination of the three predictors groups (TC-CH-PH), for each index (FEve = functional evenness, FRic = functional richness, CWM = community weighted mean, Min = minimum, Max = maximum) and each trait (H = height, SLA =specific leaf area, LDMC = leaf dry matter content, LNC = leaf nitrogen content, SM = seed mass).

		R^2				Correlation			
		TC	**TC-CH**	**TC-PH**	**TC-CH-PH**	**TC**	**TC-CH**	**TC-PH**	**TC-CH-PH**
log H	**Rao**	0.139	0.211	0.191	0.252	0.063	-0.009	0.018	0.000
	FEve	0.090	0.135	0.128	0.176	-0.006	-0.027	-0.020	0.005
	FRic	0.170	0.226	0.227	0.256	0.123	0.112	0.131	0.073
	CWM	**0.558**	**0.622**	**0.608**	**0.674**	**0.672**	**0.642**	**0.675**	**0.656**
	5^{th} qu	**0.666**	**0.698**	**0.713**	**0.743**	**0.762**	**0.746**	**0.779**	**0.750**
	95^{th} qu	**0.643**	**0.714**	**0.721**	**0.750**	**0.759**	**0.761**	**0.787**	**0.774**
log SLA	**Rao**	0.437	0.548	0.480	0.556	0.548	0.617	0.559	0.592
	FEve	0.128	0.191	0.184	0.245	0.064	0.049	0.099	0.078
	FRic	0.486	**0.595**	**0.556**	**0.606**	**0.598**	**0.653**	**0.639**	**0.657**
	CWM	**0.525**	**0.664**	**0.565**	**0.682**	**0.617**	**0.701**	**0.621**	**0.678**
	5^{th} qu	0.287	0.394	0.420	0.456	0.388	0.413	0.476	0.435
	95^{th} qu	**0.720**	**0.785**	**0.735**	**0.794**	**0.804**	**0.829**	**0.805**	**0.818**
log LDMC	**Rao**	0.194	0.370	0.244	0.396	0.197	0.359	0.224	0.341
	FEve	0.151	0.224	0.174	0.256	0.135	0.154	0.086	0.170
	FRic	0.150	0.218	0.219	0.262	0.174	0.221	0.210	0.227
	CWM	0.214	0.389	0.282	0.442	0.247	0.365	0.238	0.366
	5^{th} qu	0.146	0.266	0.226	0.301	0.068	0.166	0.148	0.148
	95^{th} qu	0.487	**0.639**	**0.514**	**0.662**	**0.617**	**0.670**	**0.576**	**0.672**
log N	**Rao**	0.247	0.354	0.320	0.389	0.306	0.298	0.322	0.299
	FEve	0.079	0.124	0.092	0.146	-0.004	0.000	-0.054	-0.031
	FRic	0.284	0.409	0.422	0.469	0.344	0.398	0.423	0.387
	CWM	0.310	0.486	0.369	**0.524**	0.384	0.492	0.359	0.480
	5^{th} qu	0.387	0.468	0.441	**0.501**	0.491	**0.504**	0.480	0.478
	95^{th} qu	0.142	0.297	0.351	0.391	0.043	0.231	0.290	0.282
log SM	**Rao**	0.182	0.309	0.219	0.330	0.214	0.323	0.187	0.265
	FEve	0.114	0.254	0.220	0.310	0.008	0.091	0.040	0.078
	FRic	0.206	0.405	0.243	0.424	0.237	0.421	0.218	0.368
	CWM	0.277	**0.506**	**0.524**	**0.605**	0.345	0.482	**0.520**	**0.548**
	5^{th} qu	0.319	0.408	0.438	0.466	0.405	0.395	0.465	0.399
	95^{th} qu	0.218	0.419	0.303	0.444	0.179	0.367	0.196	0.352

Ranking of environmental variable

The variation partitioning analysis (Fig. 2, Supplementary material Appendix 1, Table A4) showed that topographic and climatic predictors explained the greatest fraction of the variance across all models, accounting for 23% of the total variance on average (Fig. 2). The chemical edaphic predictors were important in some cases, primarily in the LDMC indices models, and explained on average 8.6% of variance for all indices and all traits. The physical edaphic predictors alone explained little variance (4.3% for all of the indices and traits on average), but the proportion of variance explained by the physical predictors was important when combined with the chemical edaphic group, especially among the LNC and SM indices.

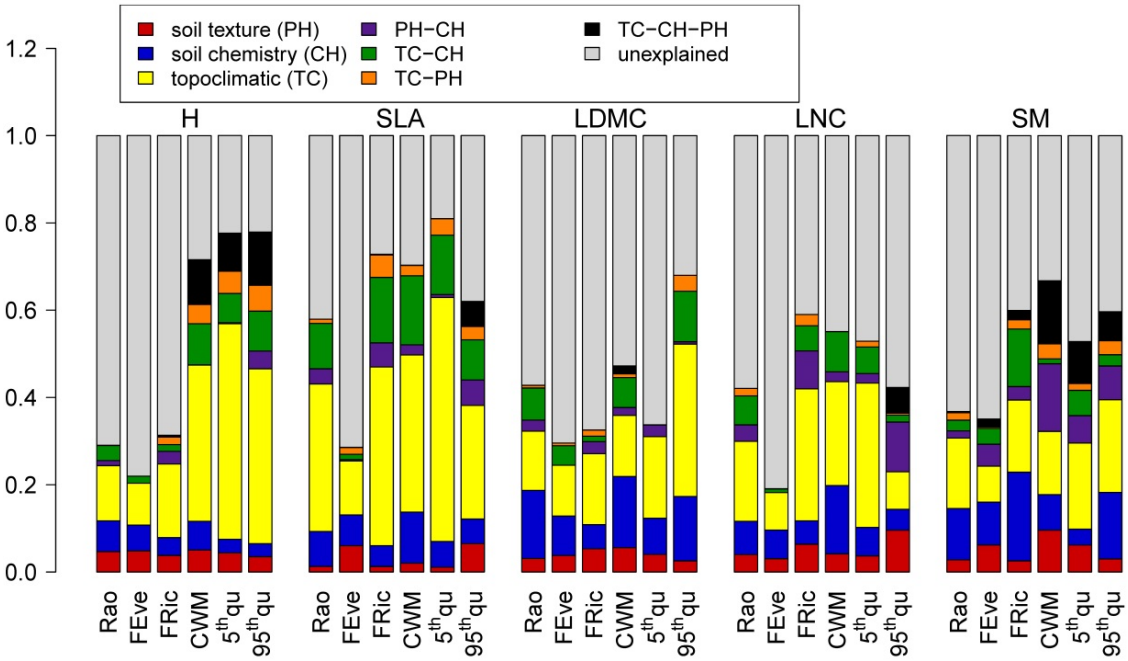


Figure 2. Barplot showing the proportion of variance - of the different indices depicting community-level patterns of traits - explained by each group of predictors and each combination of these groups. These proportions were calculated with a variance partitioning analysis. The grey portion of the bar is the proportion of the variance unexplained by the predictors used in the model. SLA = specific leaf area, H = vegetative height, LDMC = leaf dry matter content, LNC = leaf dry matter content, SM = seed mass, CWM = community weighted mean, FRic = functional richness, FEve= functional evenness.

Projections under current and future climatic conditions

The fit and evaluation results (R^2 and Spearman correlations) for the models using only topo-climatic predictors are summarised in Supplementary material Appendix 1, Table A5. The distribution of their residuals is presented in Supplementary material Appendix 1, Fig. A2. H and SLA were the traits for which model evaluations were better on average. For this reason and because these two traits express two different axes of functional differentiation, hereafter we only discuss projections for these traits (H and SLA) and metrics (CWM, 5^{th} quantile, 95^{th} quantile and FRic). We chose not to project the other functional diversity indices (FEve and Rao) in warmer conditions as they showed poorer evaluation metrics. Figure 3 shows the projections of CWM and FRic for H and SLA, displaying the decrease in vegetation height and SLA as well as trait diversity as the elevation increase and environmental condition become harsher.

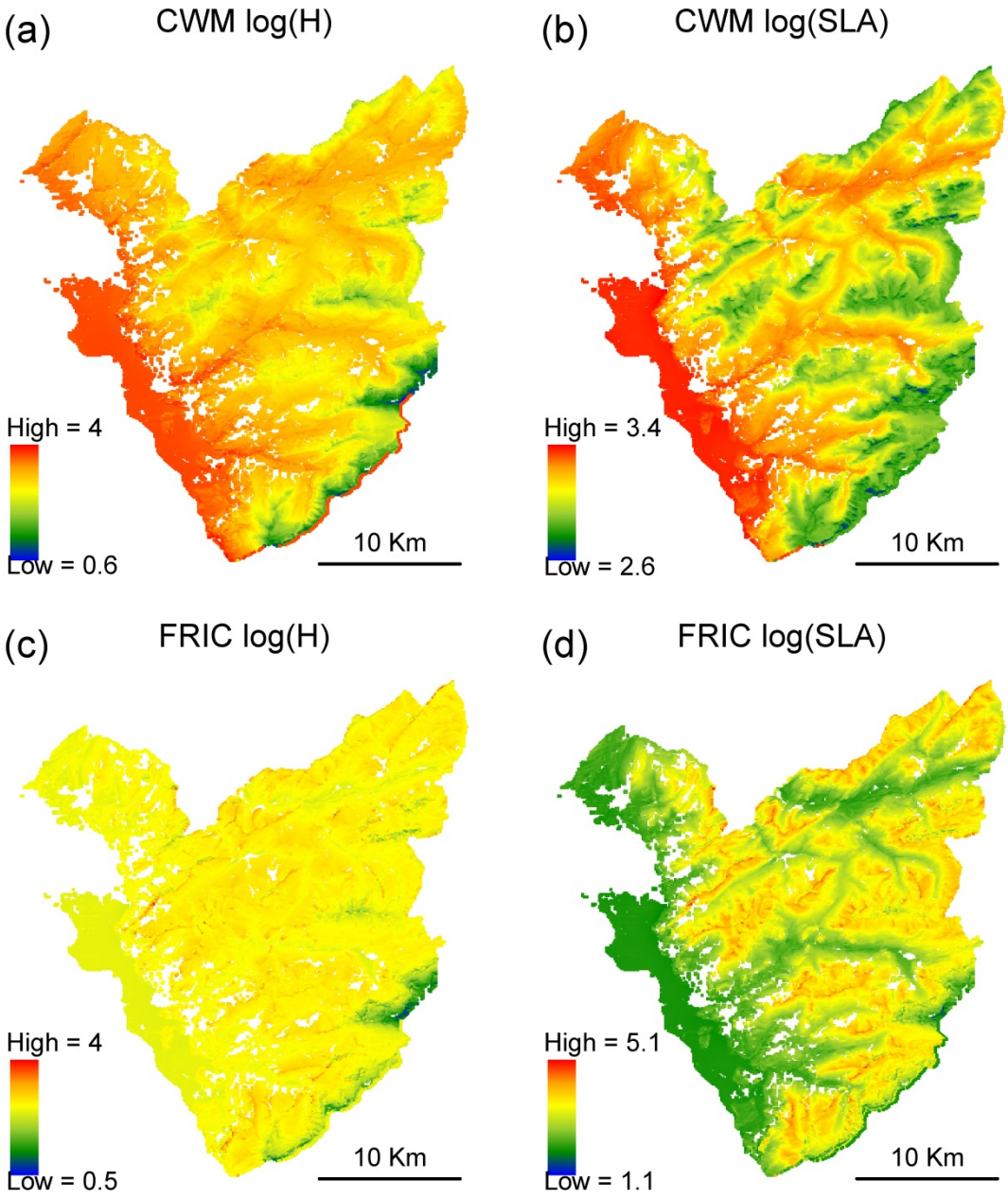


Figure 3. Model projections for the (a, b) community weighted mean (CWM) and (c, d) functional richness (FRic) for (a, c) H (m) and (b, d) SLA (mm^2*g^{-1}), for current climatic conditions on the study area (Swiss western Alps). White areas represent forested or urbanised areas that were discarded from the projections.

Under warmer climatic conditions, our models project important changes in trait values along elevation gradient (Fig. 4). Projections for H and SLA both show a global increase for those trait values especially above 1500 m (Fig. 4 a to l), followed by a decrease at the highest elevation for 5^{th} quantile of SLA (Fig. 4 g, h). Functional richness for H is predicted to increase with elevation (Fig. 4 m, n), while SLA is predicted to decrease in richness between 1000 and 2000 m and to increase above 2000 m (Fig. 4 o, p). These trends are consistent between projections for the years 2040 and 2070, though those for 2040 are weaker. Finally, Fig. 4 shows that the results for areas under 1000 m elevation in the 2040 projection and under 1500 m in the 2070 projection should be considered with caution because the climates predicted under these elevations would differ significantly from the climate used for the model calibration (*i.e.*, non-analoguous situations represented in grey). The results for the A1FI scenario are presented in the supplementary material (Supplementary material Appendix 1, Fig. A1).

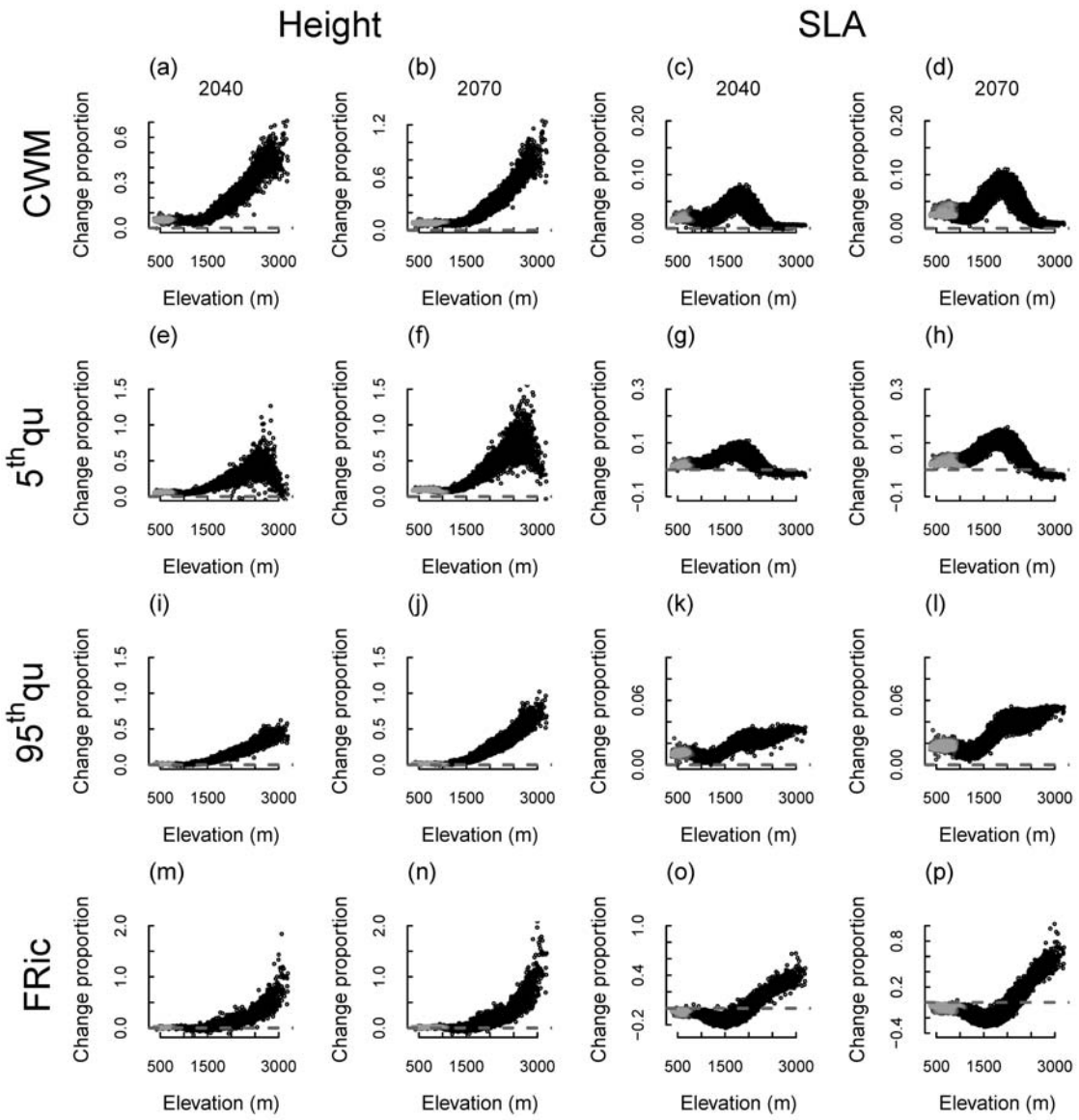


Figure 4. The proportions of change in community weighted mean (CWM), 5th quantile, 95th quantile and functional richness (FRic) of specific leaf are (SLA) and vegetative height (H) predicted between present conditions and 2040 and current conditions and 2070 according to the B2 climate change scenario. Data apply to all pixels of the study area and are plotted along elevation. The portion of the study area in which the predicted climate does not currently occur is represented in grey; the dashed line indicates the absence of change.

DISCUSSION

Our study demonstrates that building statistical models of the functional attributes of plant communities, particularly for extreme values and mean of trait distribution, can provide valuable insights into the abiotic drivers of communities' assembly. Second, we showed that useful spatial projections of community functional properties can be obtained from successful models for some traits and indices under present and future conditions. In the mountainous landscape that characterises our study area, topographic and climatic factors are more important predictors of community-level patterns of plant functional traits than soil factors. This enables the models to be projected in the future to assess climate-induced changes of functional patterns and to identify areas most susceptible to functional changes in the future.

Varying prediction accuracy among functional indices

Not all the investigated descriptors of community-level patterns of functional traits could be predicted with good accuracy. In this regard, our results revealed that the CWM and extreme (quantiles) values for SLA and H could be modelled accurately, whereas models of those indices for LDMC, LNC and SM showed lower performances. Sonnier *et al*. (2010a) attempted to predict community weighted mean and variance for several traits, including the five used here, along stress and disturbance gradients and also obtained higher correlation coefficient for CWM-H.

Habitat filtering is expected to restrict the range of suitable trait values allowing species to grow, survive and reproduce in given environmental conditions. Such restriction may preferentially operate on the lower and/or upper tail of the trait distribution of the assembled species. In this respect, extreme values of trait patterns at the community level are of high interest to derive trait-based assembly rules. Surprisingly, studies interested in species assembly have neglected this aspect of trait patterns at the community level. Beyond their ecological meaning, our results suggest that extreme values of trait distribution can be predicted in space with fairly good accuracy.

Results for the diversity indices were unequal. The model for FRic (reflecting the size of the functional space) obtained the best results among models for diversity indices. This index has been shown to correlate well with species richness (Mouchet et al. 2010, Schleuter et al. 2010), a trend that was confirmed by our study. We could thus expect the good spatial predictions observed for FRic, because species richness had already been successfully predicted in our study area (Dubuis et al. 2011). In contrast, the

models for FEve and Rao that are less correlated with species richness, showed weaker fit and predictive accuracy. This suggests that indices based on the distribution of trait abundances inside the functional space, such as FEve and Rao, correlate weakly with abiotic topo-climatic factors, at least with those used in this study. This is consistent with findings from Swenson and Weiser (2010) which could evidence only weak correlations between tree functional traits variance and climatic variables. Predictors reflecting environmental stochasticity and heterogeneity, as disturbance, land use or biotic interactions could be more suitable for predicting these indices, as relationships between functional diversity and these gradients have already been established (Grime 2006, Laliberte et al. 2010, Mayfield et al. 2010). Unfortunately, these variables were not available as spatially-explicit predictors, although they are certainly important considering the significant human impact in our study area (Randin et al. 2009).

Variables affecting community trait response

The importance and influence of predictors on the community functional patterns depended on the traits under study. For all traits and indices, topo-climatic factors explained the largest part of the variation, a consistent result with the study of Pakeman et al. (2009). But the group of topo-climatic factors was also containing the five predictors while the two other groups were containing three each. Topo-climatic factors were particularly important for SLA and H, two traits strongly related to temperature (Körner et al. 1989, Diaz et al. 1998, Körner 2003). Under current climate, the mean H and SLA values of the community decreased with decreasing temperature, as well as increasing values of topography (indicating ridges and tops). This indicates that vegetation tends to be shorter with more resistant leaves in harsh conditions (low temperature and high exposure to wind). SLA values also indicate how plants acquire, conserve and release resources (Cornelissen et al. 2003). In cold conditions, resources are scarce, and mainly species with low SLA are able to conserve (rather than exploit) resources in their leaves and tolerate those conditions.

Trends found for LDMC (increase of community mean in cold environment and on ridges) and LNC (decrease of community mean in cold environment) reflect this shift in strategy, from more exploitative to more conservative species towards higher elevations. CWM SM also decreases with decreasing temperature reflecting a trade-off in resources allocated to sexual reproduction, seeds have been shown to become smaller but more numerous at higher elevation (Pellissier et al. 2010a, Dainese and Sitzia 2013). LDMC, LNC and SM, could also be expected to correlate with soil

predictors that alter plant nutrition either directly, as in the case of nitrogen or phosphorus content, or indirectly through pH and soil texture influencing nutrient and water availability (Gobat et al. 2010). The important part of variance explained by soil predictors in the models for LDMC, LNC and SM confirmed this influence and matched the outcomes of other studies (Garnier et al. 2004, Pakeman et al. 2009, Dainese and Sitzia 2013). Hence, the projection of these traits models in space and time would be unreliable without including spatially-explicit soil predictors.

Current and future projections to assess changes in functional trait structure

By projecting our models under climate change scenarios, we expected to see those functional patterns observed at low elevation replacing those observed at higher elevation (Randin et al. 2010). Our climate change projections showed that, between 1000 and 2000 m, the strong environmental filtering caused by the current climatic conditions (which are relatively stressful for plant growth) would be relaxed, allowing higher trait means for both H and SLA and higher 5^{th} and 95^{th} quantile values. However, looking at projections of the future functional richness, we observed a decrease for SLA in this intermediate elevation belt. This suggests a homogenisation of communities from a functional point of view. Above 2000 m, H values are still predicted to increase while changes for SLA are predicted to be less important (increase of the 95^{th} quantile and decrease of the 5^{th} quantile, leading to approximately no change on average).

Cautionary note and future perspective

We must remain cautious when interpreting the patterns for the lower end of the elevational gradient (the grey pixels in Fig. 4). First, the future climate predicted to occur in these altitudes differs completely from the current climate in the study area. As a consequence, these new climatic conditions have not been taken into account during model calibration. It is very probable that species possessing new functional strategies will colonise these low elevation areas, and this could totally change the community mean traits and cause either an increase in the functional diversity of the community or, in the case of colonisation by strong competitors, lead to a decrease. Neither of these two opposing hypotheses can be favoured with our empirical approach and both could be considered equally probable. Second, because the models used for projections contained only topo-climatic predictors, the edaphic factors that were significant for some traits (especially LDMC and LNC) are omitted. This may explain the low

evaluation values of topo-climatic models for these traits. Soil moisture that could be highly important in this context is also missing. Land use and socio-economic components could also improve predictions of functional patterns in a study area strongly affected by human influences (Laliberte et al. 2010, Lavorel et al. 2011, Pakeman 2011). Moreover, for greater tractability, our study focused only on open non-woody ecosystems and did not address forested areas. As a consequence, our results did not account for possible scenarios of forest recolonisation, which have been shown to occur (Pellissier et al. 2012) and could drastically change the observed functional patterns.

To summarize, we demonstrated that the mean, 5^{th} and 95^{th} quantiles of key functional traits - associated with plant growth, persistence and reproduction - within communities, can be modelled and projected in space. Such models provide interesting tools to study community assembly as a function of abiotic environmental condition. In contrast, more research would be needed to understand which environmental factors can help predicting functional diversity indices. We showed that topo-climatic factors were more important than the retained edaphic predictors in explaining functional indices, and that the respective importance of the predictors varied according to the functional trait under study. Finally, model projections under climate warming scenarios reported important changes in community-level patterns of traits (increase in mean vegetation height, and mean SLA but decrease of SLA richness at middle elevation) more particularly from middle to highest elevations of the study area.

ACKNOWLEDGEMENTS

We are grateful to all those who helped with the field-work, especially F. Bonnet, P. Châtelain, A. Coiana and D. Paulin for helping to collect traits data, as well as S. Giovanettina for assistance with the soil analysis. We thank B. Petitpierre for providing the maps of future climates, G. Litsios and two reviewers for helpful comments on the manuscript. This study received support from the Swiss National Fund for research (SNF grant Nr. 31003A-125145, BIOASSEMBLE project) and from the European Commission (ECOCHANGE Project).

REFERENCES

Ackerly, D. D. and Cornwell, W. K. 2007. A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. — Ecol. Lett. 10: 135-145.

Borcard, D. et al. 1992. Partialling out the spatial component of ecological variation. — Ecology 73: 1045-1055.

Cornelissen, J. H. C. et al. 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide —Aust. J. Bot. 51: 335-380

Cornwell, W. K. and Ackerly, D. D. 2009. Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. — Ecol. Monogr. 79: 109-126.

Dainese, M. and Sitzia, T. 2013. Assessing the influence of environmental gradients on seed mass variation in mountain grasslands using a spatial phylogenetic filtering approach. — Perspect. Plant Ecol. Evol. Syst. 15: 12-19.

de Bello, F. et al. 2009. Partitioning of functional diversity reveals the scale and extent of trait convergence and divergence. — J. Veg. Sci. 20: 475-486.

de Bello, F. et al. 2010. Towards an assessment of multiple ecosystem processes and services via functional traits. — Biodivers. Conserv. 19: 2873-2893.

Diaz, S. et al. 1998. Plant functional traits and environmental filters at a regional scale. — J. Veg. Sci. 9: 113-122.

Diaz, S. and Cabido, M. 2001. Vive la difference: plant functional diversity matters to ecosystem processes. — Trends Ecol. Evol. 16: 646-655.

Douma, J. C. et al. 2012. Towards a functional basis for predicting vegetation patterns; incorporating plant traits in habitat distribution models. — Ecography 35: 294-305.

Dubuis, A. et al. 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. — Divers. Distrib. 17: 1122-1131.

Dubuis, A. et al. In press. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. — J. Veg. Sci.

Garnier, E. et al. 2004. Plant functional markers capture ecosystem properties during secondary succession. — Ecology 85: 2630-2637.

Gobat, J.-M. et al. 2010. Le sol vivant. — Presses polytechniques et universitaires romandes.

Grime, J. P. 1998. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. — J. Ecol. 86: 902-910.

Grime, J. P. 2006. Trait convergence and trait divergence in herbaceous plant communities: Mechanisms and consequences. — J. Veg. Sci. 17: 255-260.

Hastie, T. and Tibshirani, R. 1990. Generalized Additive Models. — Chapman and Hall.

Hector, A. et al. 1999. Plant diversity and productivity experiments in European grasslands. — Science 286: 1123-1127.

Hooper, D. U. et al. 2005. Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. — Ecol. Monogr. 75: 3-35.

Keddy, P. A. 1992a. A pragmatic approach to functional ecology. — Funct. Ecol. 6: 621-626.

Keddy, P. A. 1992b. Assembly and response rules: two goals for predictive community ecology. — J. Veg. Sci. 3: 157-164.

Körner, C. et al. 1989. Functional morphology of mountain plants. — Flora 182: 353–383.

Körner, C. 2003. Alpine Plant Life (2nd Edition). — Springer.

Küster, E. C. et al. 2010. Modelling the impact of climate and land use change on the geographical distribution of leaf anatomy in a temperate flora. — Ecography 12: 2001-2012.

Laliberte, E. and Legendre, P. 2010. A distance-based framework for measuring functional diversity from multiple traits. — Ecology 91: 299-305.

Laliberte, E. et al. 2010. Land-use intensification reduces functional redundancy and response diversity in plant communities. — Ecol. Lett. 13: 76-86.

Lavorel, S. and Garnier, E. 2002. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. — Funct. Ecol. 16: 545-556.

Lavorel, S. et al. 2011. Using plant functional traits to understand the landscape distribution of multiple ecosystem services. — J. Ecol. 99: 135-147.

Mayfield, M. M. et al. 2010. What does species richness tell us about functional trait diversity? Predictions and evidence for responses of species and functional trait diversity to land-use change. — Global Ecol. Biogeogr. 19: 423-431.

McGill, B. J. et al. 2006. Rebuilding community ecology from functional traits. — Trends Ecol. Evol. 21: 178-185.

Moles, A. T. et al. 2008. A new framework for predicting invasive plant species. — J. Ecol. 96: 13-17.

Mouchet, M. A. et al. 2010. Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. — Funct. Ecol. 24: 867-876.

Nagelkerke, N. J. D. 1991. A note on a general definition of the coefficient of determination. — Biometrika 78: 691-692.

Nakicenovic, N. and Swart, R. 2000. Special Report on Emissions Scenarios. Intergovernmental Panel on Climate Change (IPCC). — Cambridge University Press.

Nogues-Bravo, D. and Rahbek, C. 2011. Communities Under Climate Change. — Science 334: 1070-1071.

Pakeman, R. 2011. Functional diversity indices reveal the impacts of land use intensification on plant community assembly. — J. Ecol. 99: 1143-1151.

Pakeman, R. J. and Quested, H. M. 2007. Sampling plant functional traits: What proportion of the species need to be measured? — Appl. Veg. Sci. 10: 91-96.

Pakeman, R. J. et al. 2009. Relative climatic, edaphic and management controls of plant functional trait signatures. — J. Veg. Sci. 20: 148-159.

Pellissier, L. et al. 2010a. Plant traits co-vary with altitude in grasslands and forests in the European Alps. — Plant Ecol. 211: 351-365.

Pellissier, L. et al. 2010b. Spatial pattern of floral morphology: possible insight into the effects of pollinators on plant distributions. — Oikos 119: 1805-1813.

Pellissier, L. et al. 2012. Spatial predictions of land use transitions and associated threats to biodiversity: the case of forest regrowth in mountain grasslands. — Appl. Veg. Sci. in press:

Randin, C. F. et al. 2009. Land use improves spatial predictions of mountain plant abundance but not presence-absence. — J. Veg. Sci. 20: 996-1008.

Randin, C. F. et al. 2010. Using georeferenced databases to assess the effect of climate change on alpine plant species and diversity. — In: Körner, C. S., E. (ed), Data mining for global trends in mountain biodiversity. CRC/Taylor & Francis.

Schleuter, D. et al. 2010. A user's guide to functional diversity indices. — Ecol. Monogr. 80: 469-484.

Shipley, B. et al. 2006. From plant traits to plant communities: A statistical mechanistic approach to biodiversity. — Science 314: 812-814.

Sonnier, G. et al. 2010a. Plant traits, species pools and the prediction of relative abundance in plant communities: a maximum entropy approach. — J. Veg. Sci. 21: 318-331.

Sonnier, G. et al. 2010b. Quantifying relationships between traits and explicitly measured gradients of stress and disturbance in early successional plant communities. — J. Veg. Sci. 21: 1014-1024.

Swenson, N. G. and Weiser, M. D. 2010. Plant geography upon the basis of functional traits: an example from eastern North American trees. — Ecology 91: 2234-2241.

Swenson, N. G. et al. 2012. The biogeography and filtering of woody plant functional diversity in North and South America. — Global Ecol. Biogeogr. 21: 798-808.

Symstad, A. J. 2000. A test of the effects of functional group richness and composition on grassland invasibility. — Ecology 81: 99-109.

Tilman, D. 1997. Distinguishing between the effects of species diversity and species composition. — Oikos 80: 185-185.

Vaieretti, M. V. et al. 2007. Two measurement methods of Leaf Dry Matter Content produce similar results in a broad range of species. — Ann. Bot. 99: 955-958.

Van Bodegom, P. M. et al. 2012. Going beyond limitations of plant functional types when predicting global ecosystem–atmosphere fluxes: exploring the merits of traits-based approaches. — Global Ecol. Biogeogr. 21: 625-636.

Venn, S. E. et al. 2011. Using plant functional traits to explain community composition across a strong environmental filter in Australian alpine snowpatches. — Plant Ecol. 212: 1491-1499.

Violle, C. et al. 2007. Let the concept of trait be functional! — Oikos 116: 882-892.

Vittoz, P. and Guisan, A. 2007. How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. — Journal of Vegetation Science 18: 413-422.

Webb, C. T. et al. 2010. A structured and dynamic framework to advance traits-based theory and prediction in ecology. — Ecol. Lett. 13: 267-283.

Westoby, M. 1998. A leaf-height-seed (LHS) plant ecology strategy scheme. — Plant Soil 199: 213-227.

Wilson, P. J. et al. 1999. Specific Leaf Area and Leaf Dry Matter Content as alternative predictors of plant strategies. — New Phytol. 143: 155-162.

Zimmermann, N. E. and Kienast, F. 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. — J. Veg. Sci. 10: 469-482.

SUPPLEMENTARY INFORMATIONS

Table A1. Pearson correlation values between all topo-climatic and edaphic variables used as predictors in the models.

	pH	P	Cl	Sa	Degree day	Slope	Moisture	Radiation	Topo
N	0.05	0.43	0.41	-0.28	-0.17	-0.18	0.05	0.09	0.18
pH		0.07	-0.11	-0.13	0	0.15	0.02	0.06	-0.09
P			0.25	-0.18	-0.09	-0.23	-0.03	0.04	0.1
Cl				-0.7	0.09	-0.1	-0.2	0.23	0.02
Sa					-0.04	0.06	0.1	-0.08	0.04
Degree day						-0.26	-0.83	0.06	-0.57
Slope							0.33	-0.1	0.22
Moisture								-0.43	0.45
Radiation									0.09

Table A2. R^2 and p-values resulting from linear regressions between functional indices for each traits and each environmental predictor.

		degree-days		moisture		solar radiation		slope		topographic position		N		pH		P		Cl		Sa	
		R2	p-val	R2	p-val	R2	p-val	R2	p-val	R2	p-val	R2	p-val	R2	p-val	R2	p-val	R2	p-val	R2	p-val
log H	**Rao**	0.021	0.133	0.012	0.308	0.011	0.351	0.040	0.021	0.018	0.170	0.006	0.574	0.011	0.358	0.029	0.061	0.018	0.182	0.002	0.804
	FEve	0.024	0.095	0.015	0.236	0.012	0.330	0.012	0.316	0.023	0.109	0.011	0.338	0.028	0.066	0.003	0.743	0.013	0.274	0.015	0.242
	FRic	0.064	0.002	0.035	0.032	0.002	0.801	0.031	0.047	0.010	0.393	0.019	0.160	0.025	0.085	0.025	0.086	0.017	0.189	0.006	0.561
	CWM	0.412	0.000	0.352	0.000	0.026	0.084	0.005	0.649	0.107	0.000	0.036	0.030	0.100	0.000	0.005	0.609	0.129	0.000	0.030	0.057
	5th qu	0.598	0.000	0.515	0.000	0.033	0.042	0.023	0.110	0.184	0.000	0.010	0.385	0.070	0.001	0.005	0.609	0.112	0.000	0.040	0.019
	95th qu	0.576	0.000	0.439	0.000	0.010	0.368	0.005	0.602	0.149	0.000	0.020	0.144	0.126	0.000	0.000	0.988	0.175	0.000	0.044	0.013
log SLA	**Rao**	0.266	0.000	0.193	0.000	0.007	0.520	0.015	0.244	0.083	0.000	0.038	0.026	0.100	0.000	0.003	0.748	0.023	0.111	0.001	0.883
	FEve	0.015	0.244	0.011	0.341	0.007	0.494	0.030	0.052	0.046	0.011	0.020	0.146	0.008	0.466	0.005	0.594	0.002	0.810	0.012	0.312
	FRic	0.380	0.000	0.282	0.000	0.004	0.687	0.049	0.008	0.097	0.000	0.059	0.003	0.029	0.059	0.004	0.681	0.026	0.079	0.006	0.580
	CWM	0.435	0.000	0.316	0.000	0.003	0.783	0.033	0.042	0.210	0.000	0.048	0.009	0.107	0.000	0.003	0.784	0.013	0.279	0.012	0.318
	5th qu	0.648	0.000	0.527	0.000	0.019	0.157	0.058	0.003	0.218	0.000	0.029	0.060	0.047	0.010	0.002	0.849	0.016	0.213	0.012	0.317
	95th qu	0.353	0.000	0.223	0.000	0.003	0.749	0.008	0.456	0.134	0.000	0.009	0.440	0.115	0.000	0.004	0.712	0.176	0.000	0.022	0.115
log LDMC	**Rao**	0.093	0.000	0.057	0.004	0.011	0.347	0.004	0.654	0.041	0.018	0.029	0.058	0.092	0.000	0.006	0.588	0.024	0.099	0.010	0.388
	FEve	0.030	0.053	0.012	0.321	0.039	0.023	0.032	0.042	0.017	0.190	0.029	0.061	0.031	0.047	0.002	0.806	0.001	0.923	0.017	0.197
	FRic	0.078	0.000	0.061	0.002	0.017	0.199	0.008	0.480	0.051	0.007	0.003	0.737	0.064	0.002	0.001	0.890	0.012	0.313	0.033	0.039
	CWM	0.034	0.035	0.038	0.024	0.022	0.121	0.055	0.004	0.079	0.000	0.046	0.012	0.030	0.056	0.023	0.106	0.067	0.001	0.009	0.413
	5th qu	0.008	0.471	0.028	0.066	0.018	0.176	0.016	0.207	0.024	0.095	0.042	0.017	0.015	0.234	0.005	0.651	0.032	0.042	0.001	0.890
	95th qu	0.391	0.000	0.312	0.000	0.016	0.218	0.049	0.008	0.234	0.000	0.031	0.050	0.097	0.000	0.022	0.123	0.019	0.158	0.004	0.685
log LNC	**Rao**	0.093	0.000	0.049	0.008	0.022	0.124	0.014	0.270	0.045	0.012	0.038	0.025	0.025	0.086	0.002	0.811	0.046	0.011	0.010	0.365
	FEve	0.017	0.187	0.005	0.647	0.020	0.138	0.002	0.793	0.015	0.243	0.016	0.211	0.004	0.654	0.009	0.402	0.000	0.993	0.003	0.742
	FRic	0.186	0.000	0.124	0.000	0.002	0.869	0.024	0.096	0.118	0.000	0.029	0.060	0.005	0.631	0.010	0.379	0.027	0.071	0.014	0.253
	CWM	0.204	0.000	0.180	0.000	0.006	0.547	0.019	0.165	0.123	0.000	0.028	0.064	0.111	0.000	0.003	0.773	0.000	0.979	0.019	0.153
	5th qu	0.229	0.000	0.188	0.000	0.003	0.784	0.062	0.002	0.113	0.000	0.006	0.542	0.037	0.028	0.002	0.820	0.005	0.620	0.015	0.229
	95th qu	0.003	0.762	0.019	0.153	0.032	0.045	0.003	0.752	0.015	0.238	0.064	0.002	0.087	0.000	0.042	0.016	0.153	0.000	0.042	0.016
log SM	**Rao**	0.025	0.108	0.025	0.109	0.001	0.951	0.060	0.005	0.004	0.715	0.019	0.182	0.028	0.083	0.027	0.093	0.024	0.118	0.007	0.560
	FEve	0.034	0.050	0.028	0.080	0.018	0.201	0.017	0.215	0.010	0.434	0.061	0.004	0.074	0.001	0.004	0.684	0.067	0.002	0.009	0.439
	FRic	0.016	0.235	0.028	0.082	0.005	0.654	0.064	0.003	0.052	0.010	0.031	0.064	0.064	0.003	0.053	0.009	0.041	0.025	0.004	0.708
	CWM	0.135	0.000	0.178	0.000	0.038	0.033	0.005	0.664	0.034	0.048	0.120	0.000	0.081	0.001	0.018	0.214	0.294	0.000	0.011	0.365
	5th qu	0.243	0.000	0.218	0.000	0.006	0.578	0.066	0.003	0.051	0.010	0.040	0.029	0.115	0.000	0.009	0.465	0.184	0.000	0.017	0.230
	95th qu	0.018	0.208	0.011	0.388	0.021	0.150	0.044	0.019	0.023	0.132	0.066	0.003	0.031	0.062	0.015	0.271	0.108	0.000	0.004	0.690

Table A3: Correlation coefficients between the functional indices for all traits and species richness

	log (H)	log (SLA)	log (LDMC)	log (LNC)	log (SM)
Rao	0.264	0.316	0.219	0.258	0.051
FEve	-0.117	-0.063	0.007	0.002	-0.267
FRic	0.564	0.471	0.368	0.403	0.517
CWM	0.403	0.120	0.269	-0.003	0.447
5^{th} qu	0.165	-0.008	0.019	-0.176	0.145
95^{th} qu	0.452	0.383	0.146	0.412	0.394

Table A4: Results of the variance partitioning analysis for each group of predictors and each combination of these groups predictors (TC = topo-climatic, CH = chemical edaphic and PH = physical edaphic), for the models for each index (FEve = functional evenness, FRic = functional richness, CWM = community wheighted mean, 5^{th} qu. = 5^{th} quantile, 95^{th} qu. = 95^{th} quantile) and each trait (H = height, SLA =specific leaf area, LDMC = leaf dry matter content, LNC = leaf nitrogen content, SM=seed mass).

		PH	CH	CL	CH+PH	CL+CH	CL+PH	CL+CH+PH	Unexplained
log (H)	**Rao**	0.047	0.071	0.126	0.012	0.034	0.014	-0.013	0.710
	FEve	0.048	0.059	0.098	-0.002	0.020	0.003	-0.007	0.781
	FRic	0.038	0.041	0.169	0.029	0.015	0.017	0.004	0.687
	CWM	0.050	0.066	0.358	0.001	0.094	0.044	0.103	0.284
	5^{th} qu	0.044	0.031	0.494	0.003	0.066	0.051	0.087	0.223
	95^{th} qu	0.035	0.030	0.401	0.041	0.091	0.059	0.122	0.221
log (SLA)	**Rao**	0.013	0.080	0.338	0.035	0.104	0.021	-0.011	0.420
	FEve	0.060	0.071	0.123	0.003	0.012	0.015	0.000	0.715
	FRic	0.013	0.048	0.410	0.055	0.150	0.052	0.002	0.272
	CWM	0.020	0.117	0.360	0.023	0.158	0.024	0.001	0.297
	5^{th} qu	0.011	0.059	0.559	0.007	0.136	0.071	-0.033	0.190
	95^{th} qu	0.065	0.056	0.260	0.058	0.092	0.031	0.057	0.380
log (LDMC)	**Rao**	0.031	0.156	0.136	0.025	0.074	0.014	-0.008	0.572
	FEve	0.038	0.090	0.123	-0.006	0.045	0.010	-0.003	0.705
	FRic	0.053	0.055	0.163	0.028	0.013	0.017	-0.003	0.675
	CWM	0.056	0.163	0.140	0.018	0.069	0.008	0.018	0.528
	5^{th} qu	0.040	0.083	0.187	0.046	-0.002	-0.018	0.000	0.663
	95^{th} qu	0.025	0.148	0.349	0.005	0.116	0.053	-0.016	0.320
log (LNC)	**Rao**	0.040	0.076	0.183	0.038	0.066	0.024	-0.007	0.579
	FEve	0.030	0.066	0.094	-0.008	0.012	-0.004	0.000	0.809
	FRic	0.064	0.054	0.302	0.087	0.057	0.060	-0.034	0.410
	CWM	0.041	0.157	0.238	0.023	0.104	-0.002	-0.009	0.449
	5^{th} qu	0.037	0.065	0.331	0.022	0.061	0.020	-0.007	0.471
	95^{th} qu	0.096	0.048	0.086	0.114	0.016	0.004	0.059	0.577
log (SM)	**Rao**	0.028	0.118	0.161	0.017	0.025	0.017	0.003	0.632
	FEve	0.062	0.098	0.082	0.050	0.036	0.003	0.018	0.649
	FRic	0.025	0.203	0.166	0.031	0.132	0.021	0.021	0.401
	CWM	0.096	0.082	0.145	0.155	0.011	0.035	0.144	0.333
	5th qu	0.062	0.036	0.197	0.063	0.058	0.016	0.096	0.472
	95^{th} qu	0.030	0.153	0.212	0.078	0.026	0.032	0.066	0.403

Table A5: R^2 and Spearman correlations between the observed and predicted functional indices values for models with topo-climatic predictors only. The lines in bold are the indices used for projections across space and time.

		R^2	Cor
log (H)	**Rao**	0.055	0.263
	FEve	0.071	0.266
	FRic	0.174	0.424
	CWM	0.677	0.820
	5^{th} qu	0.655	0.812
	95^{th} qu	0.719	0.850
log (SLA)	**Rao**	0.266	0.525
	FEve	0.065	0.264
	FRic	0.408	0.644
	CWM	0.513	0.723
	5^{th} qu	0.695	0.832
	95^{th} qu	0.506	0.719
log (LDMC)	**Rao**	0.124	0.359
	FEve	0.099	0.323
	FRic	0.086	0.310
	CWM	0.164	0.412
	5^{th} qu	0.118	0.343
	95^{th} qu	0.297	0.558

		R^2	Cor
log (N)	**Rao**	0.125	0.373
	FEve	0.109	0.346
	FRic	0.126	0.365
	CWM	0.242	0.502
	5^{th} qu	0.148	0.394
	95^{th} qu	0.221	0.481
log (SM)	**Rao**	0.042	0.223
	FEve	0.205	0.472
	FRic	0.132	0.379
	CWM	0.352	0.598
	5^{th} qu	0.298	0.546
	95^{th} qu	0.176	0.436

Figure A1

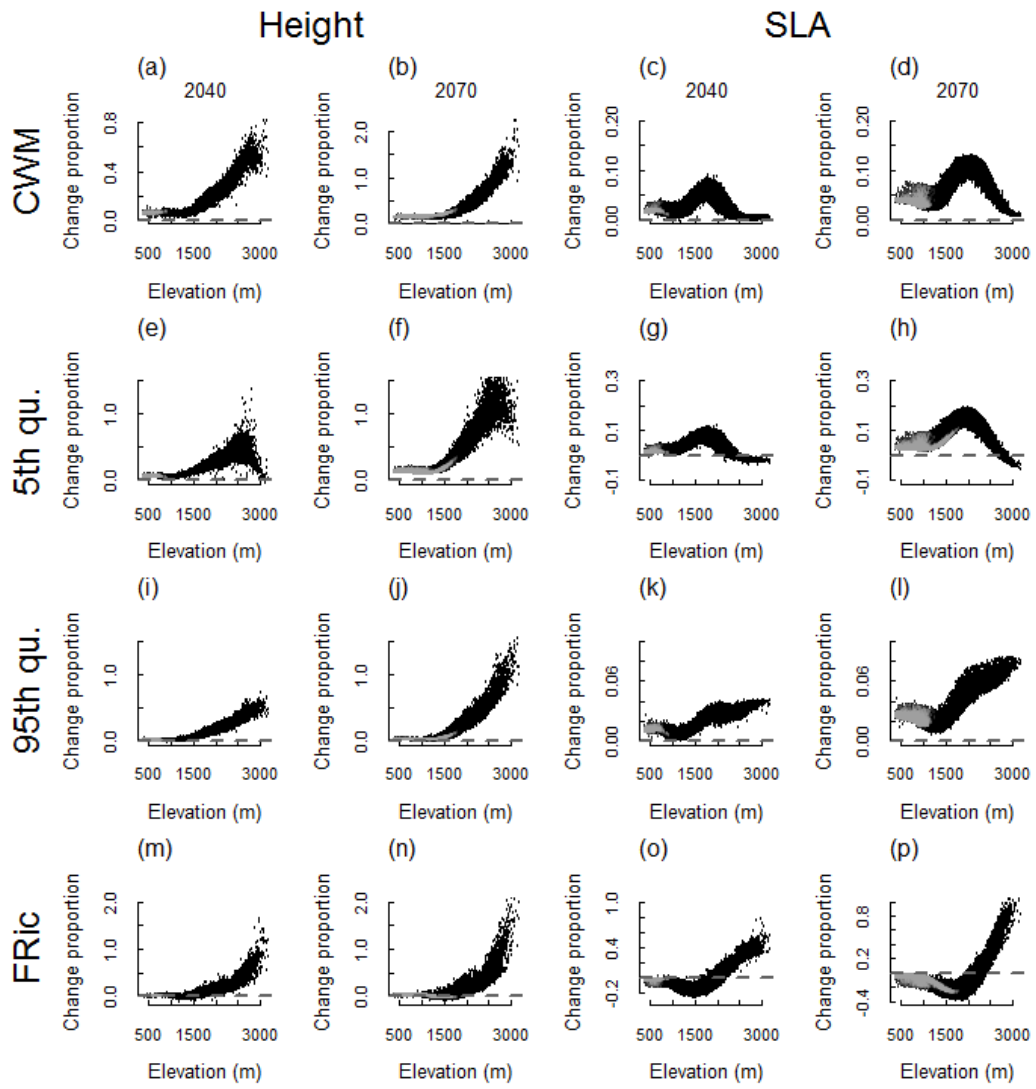


Figure A1. The proportions of change in community weighted mean (CWM), 5th and 95th quantiles and functional richness (FRic) of specific leaf are (SLA) and vegetative height (H) predicted between present conditions and 2040 and current conditions and 2070 according to the A1FI climate change scenario. Data apply to all pixels of the study area and are plotted along elevation. The portion of the study area in which the predicted climate does not currently occur is represented in grey; the dashed line indicates the absence of change.

Figure A2

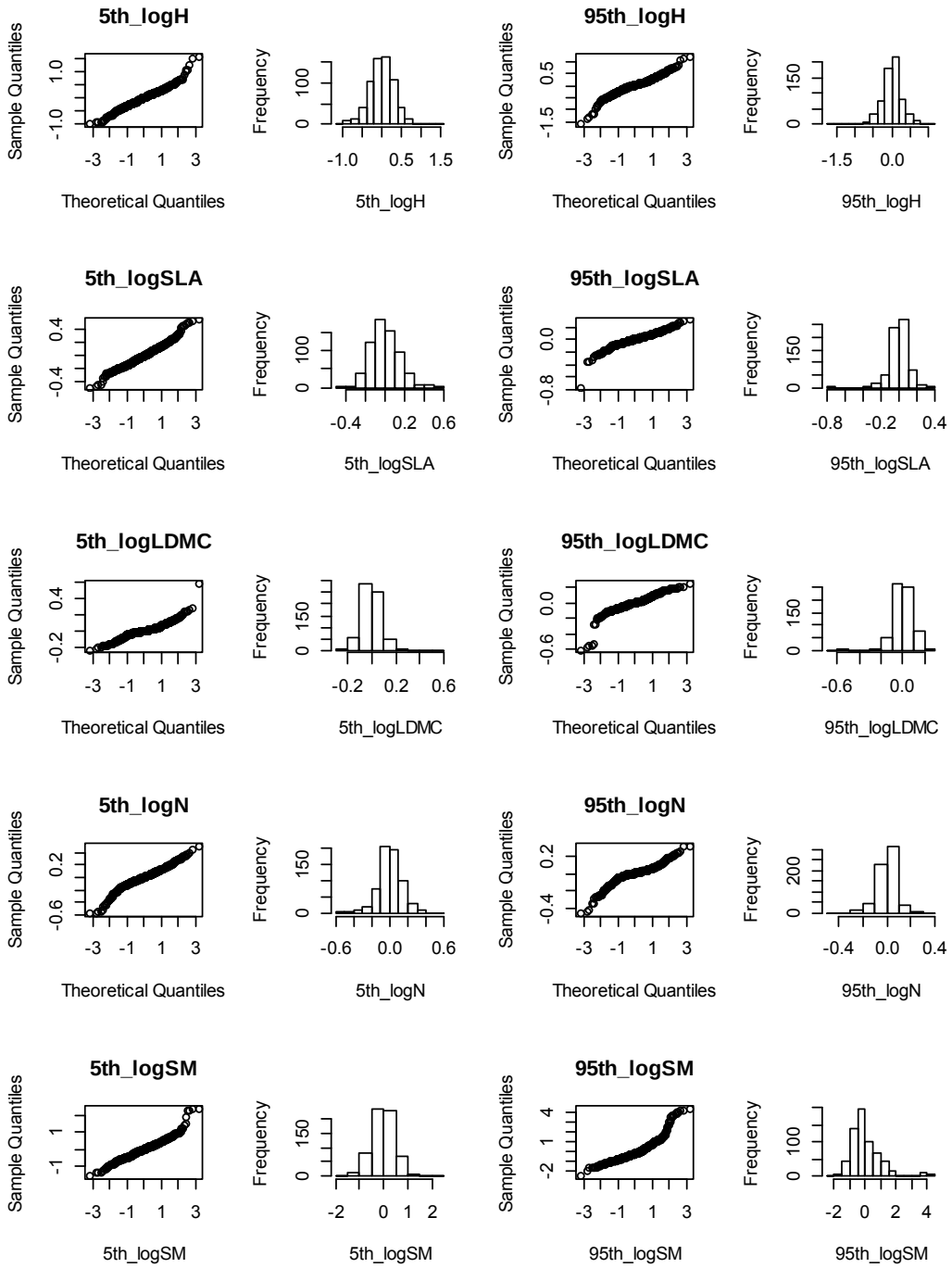

5th_logH
Sample Quantiles
Theoretical Quantiles
Frequency
5th_logH
95th_logH
Sample Quantiles
Theoretical Quantiles
Frequency
95th_logH
5th_logSLA
Sample Quantiles
Theoretical Quantiles
Frequency
5th_logSLA
95th_logSLA
Sample Quantiles
Theoretical Quantiles
Frequency
95th_logSLA
5th_logLDMC
Sample Quantiles
Theoretical Quantiles
Frequency
5th_logLDMC
95th_logLDMC
Sample Quantiles
Theoretical Quantiles
Frequency
95th_logLDMC
5th_logN
Sample Quantiles
Theoretical Quantiles
Frequency
5th_logN
95th_logN
Sample Quantiles
Theoretical Quantiles
Frequency
95th_logN
5th_logSM
Sample Quantiles
5th_logSM
Frequency
5th_logSM
95th_logSM
Sample Quantiles
95th_logSM
Frequency
95th_logSM

Figure A2 (continued)

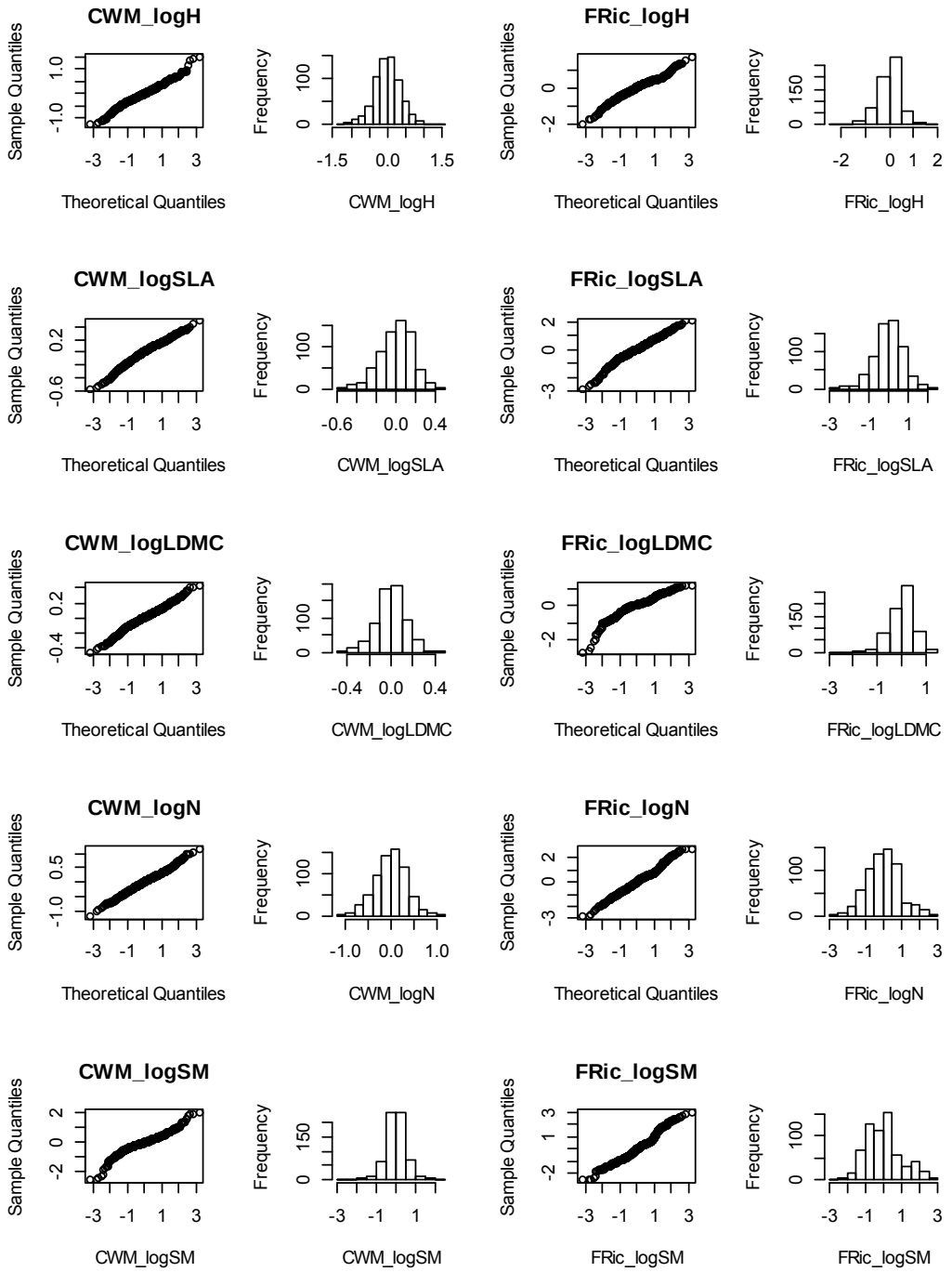


Figure A2 (continued)

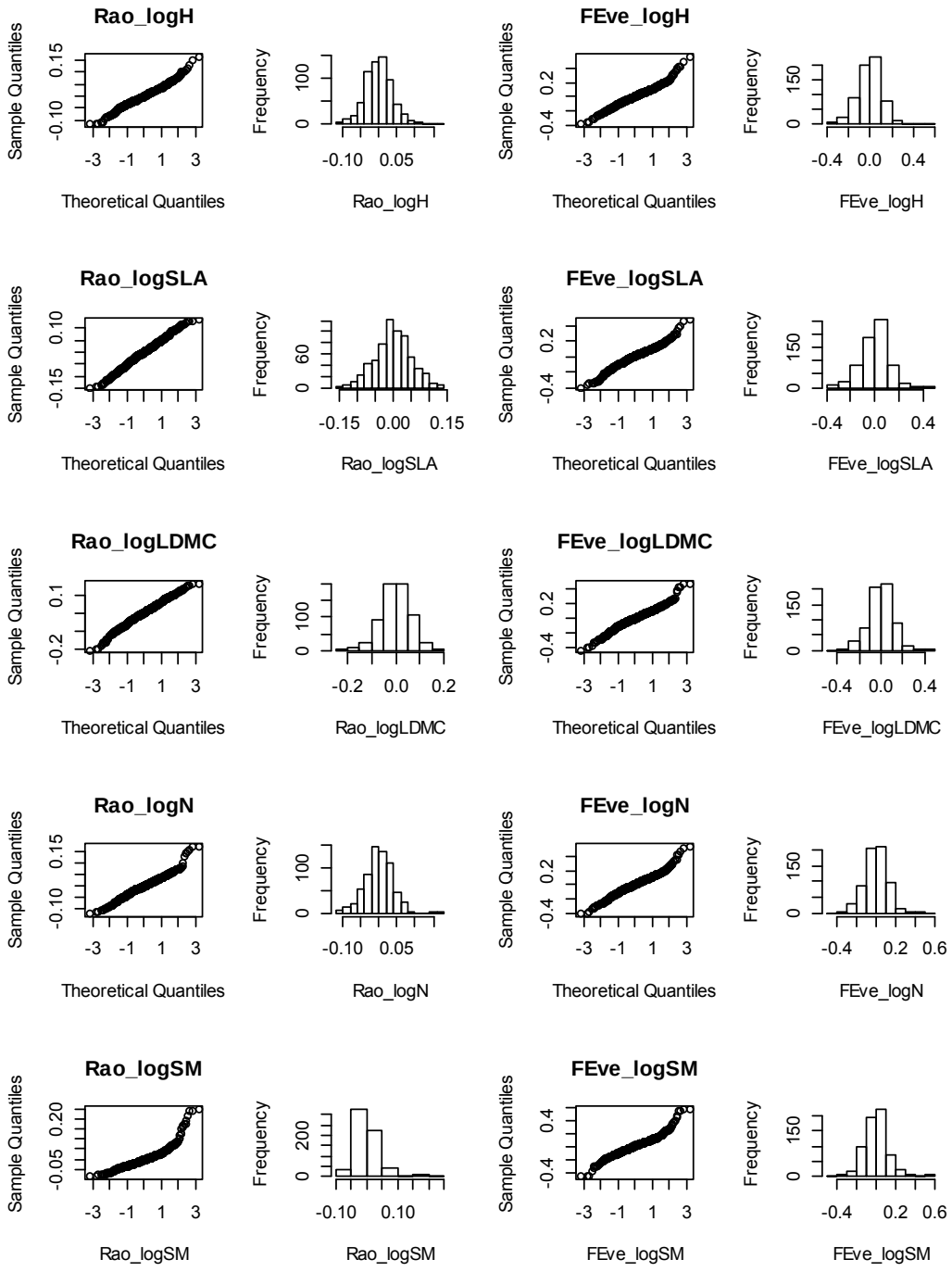


Figure A2. Quantile-quantile plots and histograms of the distribution of the residuals resulting of the model build with topo-climatic predictors and used for spatial projections, for each index (FEve = functional evenness, FRic = functional richness, CWM = community wheighted mean, 5th qu. = 5th quantile, 95th qu. = 95th quantile) and each trait (H = height, SLA =specific leaf area, LDMC = leaf dry matter content, LNC = leaf nitrogen content, SM=seed mass).

CHAPTER 5

Using macroecological modelling to constrain community predictions from species distribution models

Anne Dubuis [1]
Julien Pottier [2]
Loïc Pellissier [1]
Antoine Guisan [1,3]

1 Department of Ecology and Evolution, University of Lausanne, Biophore building, 1015 Lausanne, Switzerland

2 INRA, Grassland Ecosystem Research Unit (UREP), 5 Chemin de Beaulieu, 63100 Clermont-Ferrand, France

3 Faculty of Geoscience and the Environment, University of Lausanne, 1015 Lausanne, Switzerland

This paper is in preparation for Journal of Biogeography

Contribution to the project: I participated to the field work. I performed the statistical analysis and wrote the initial draft of the paper.

ABSTRACT

Modelling species at the assemblage level is required to make effective forecast of global change impacts on biodiversity and ecosystem functioning. Most community predictions so far have used correlative models of macroecological properties of communities (MEM), but approaches involving the stacking of individual species distribution models (S-SDMs) are increasingly used. It was recently shown that S-SDM tend to overpredict species richness. To obtain more realistic prediction of species assemblages, the SESAM framework proposed as a solution that S-SDM predictions may be further constrained by predictions from direct MEMs of community attributes (e.g., species richness, functional traits). In this paper we test this hypothesis by using a dataset of plant communities in the western Swiss Alps. Two types of macroecological constraints were tested on S-SDMs: predicted value of species richness and predicted extreme values of functional traits. We show that all constraints contribute to reduce species richness overpredictions, but that assemblage composition predictions are only slightly improved. We discuss limitation of this approach, and possible improvements for predicting composition, such as improving the causality of environmental predictors (e.g. soil properties) and testing other types of ecological or functional constraints.

Keywords

Stacked-SDM (S-SDM); SESAM framework; species distribution model (SDM); community ecology; functional ecology; macroecology; mountain system.

INTRODUCTION

Understanding the distribution and composition of species assemblages and being able to predict them in space and time are highly important tasks to investigate the fate of biodiversity in the current global change context. Different approaches have been proposed to predict the composition of species assemblages that can work on mechanistic or empirical basis. Shipley et al. (2006) proposed to use the predicted community weighted mean of functional traits to infer the assemblage composition given the traits of the species (Shipley et al., 2006, 2011; Sonnier et al., 2010a). Mokany et al. (2011, 2012) proposed another approach to relate macroecological models for species richness and composition dissimilarity to species data through a dynamic modelling framework allowing predicting assemblage composition. Another approach is to use the empirical relationships between species distribution data and environmental predictors to predict community types or axes of compositional variation derived from ordination techniques (Ferrier & Guisan, 2006). One method that is more and more used is to predict the distributions of individual species (species distribution models (SDM); based on the niche principle) and stack them to predict communities (stacked-SDM (S-SDM)). This method, also called “predict first assemble later” in Ferrier & Guisan (2006) has been tested in recent studies to draw conclusions about species richness (SR), assemblage composition or species turnover under current or future climatic conditions (Baselga & Araújo, 2009, 2010; Aranda & Lobo, 2011; Albouy et al., 2012; Pottier et al., 2012). While species models evaluated individually can yield fair to good predictions, assemblage predictions tend to show weak accuracy (Baselga & Araújo, 2010; Aranda & Lobo, 2011; Pottier et al., 2012), likely because small errors in species predictions cumulate into larger errors in assemblage predictions.

On a different issue, previous studies have also shown that S-SDM tend to overpredict, on average, species richness per unit area (Algar et al., 2009; Dubuis et al., 2011; Mateo et al., 2012). Such overprediction is however expected given that no macroecological constraint is set on the species assemblages (Guisan & Rahbek, 2011). Based on this observation, Guisan and Rahbeck (2011) proposed a framework – SESAM: spatially-explicit species assemblage modelling – that, by integrating macroecological models of assemblage properties (e.g., species richness or aggregated functional traits) and species distribution models, aims at ameliorating the prediction of species assemblage. In this context, macroecological models (MEM) are defined as models of emergent properties or attributes of communities, such as species richness (SR) or other functional characteristics that are theoretically

predictable directly from environmental variables (Dubuis et al., in review., 2011; Francis & Currie, 2003; Moser et al., 2005; Sonnier et al., 2010b). MEMs, which correspond to the "assemble first, predict later" approach of Ferrier & Guisan (2006), have been shown to provide less biased predictions of SR than S-SDMs (Dubuis et al., 2011). The main idea of the SESAM framework is to use outputs from MEMs as constraints to limit the number of species predicted with S-SDM, and this way attempt improving predictions of community composition.

SESAM comprises the following four components (Guisan & Rahbek 2011). First the species pool of each modelling unit must be defined. Secondly, SDMs are built to filter the species in the species pool according to their suitability to the environmental conditions in the modelling unit. The SDM are then stacked to represent the total, unconstrained species assemblage (S-SDM) in each unit. Thirdly, assemblage properties are predicted with one or several MEMs, defining constraints to be applied to the assemblage in each unit. Finally, the species to be kept in the assemblage are chosen among those predicted by the S-SDM, through biotic assembly rules (e.g. pairwise interactions).

In this study, we test – for the first time - a simplified version of this framework. Based on previous autocorrelation studies, we consider a unique species pool defined as the most frequent species occurring in a 700 km^2 study area in Switzerland (170 species). We test if using several ways of constraining S-SDMs with MEM predictions can improve the accuracy of SR and composition predictions. As constraints, we test the use of predicted SR and predicted extreme values of two functional traits: specific leaf area (SLA) and height (H) traits. The SR model limits the number of species occurring in each plot while the models for extreme functional characteristics limit species that can occur in each plot according to their traits, given environmental characteristics. The fourth component (biotic filtering) was not yet implemented here, as it represents a next analytical step. We thus simply selected the species according to their predicted probability of presence. The predicted assemblages are compared with the observed communities in term of species richness and composition using several evaluation metrics.

METHODS

Vegetation and traits data

The study area is located in the western Swiss Alps of Switzerland. It covers ca. 700 km^2, with elevations ranging from 375 to 3210 m and a temperate climate. The species occurrence data used in our analysis originate from fieldwork conducted between 2002 and 2009 in the study area following a random-stratified sampling design and limited to open, non-woody vegetation (see Dubuis et al. (2011) for more information). A first dataset of 613 vegetation plots of 4 m^2 each was inventoried and used for SDM and MEM calibration. An additional set of 298 plots was identically surveyed to evaluate S-SDMs, and test the MEM constraints. This evaluation dataset was shown to be spatially independent of the first one (Pottier et al., 2012).

Vegetative height (H) was measured for each species in the field as the distance between the top of the photosynthetic tissue and the ground. This trait is related to competitive ability and is correlated with above-ground biomass. Specific leaf area (SLA) was calculated as the ratio of the leaf surface to its dry mass and expressed in $mm^2\ mg^{-1}$. SLA is correlated with the relative growth rate and photosynthetic ability of herbaceous species (see Pottier et al. (2012)).

Stacked species distribution models

We modelled the distribution of 170 species (with 20 or more occurrences in the calibration dataset and available functional trait data). This set of species was considered as the study area species pool. The environmental predictors used were computed from temperature and precipitation data recorded by the Swiss network of meteorological stations and a digital elevation model at 25 meter resolution. We used growing degree days (above 0°C), moisture index over the growing season (difference between precipitation and potential evapotranspiration), the sum of solar radiations over the year, slope (in degree) and topographic position (unitless, indicating the ridges and valleys). These five variables have been shown to be useful to predict species distributions in mountainous environment (Dubuis et al., 2011). The distribution of each species was modelled using an ensemble forecasting approach based on the average of three techniques. Generalised linear models (GLM), generalised additive models (GAM), generalised boosted regression models (GBM) were fitted in R (2.14.1) with the BIOMOD library (Thuiller et al., 2009). The models were projected on the second dataset and evaluated with the area under the curve (AUC) of a receiver operating

characteristic plot (ROC). Finally the three techniques were averaged after weighting by their respective AUC. Only models with an AUC higher than 0.7 were kept in the ensemble forecast. The projected species distribution were transformed in binary presences and absences such as to maximise sensitivity and specificity (Liu et al., 2005). The resulting binary projections were stacked to predict assemblages in each of the evaluation plots (S-SDM). The steps described above were also performed with an ensemble forecasting of models evaluated and weighted with the true skill statistics (TSS; Allouche et al., 2006).

Macro-ecological models

Observed SR was calculated as the number of species (among the 170 used in this study) present in each sampling plot. It was predicted with a GAM using a Poisson distribution and smoothing function with four degrees of freedom as a function of the five predictors mentioned before. The 5^{th} and 95^{th} percentiles of log-transformed traits values were extracted for each trait and each plot. With these indices we want to reflect minimal and maximal trait values without being distorted by outliers. The two percentiles were modelled as a function of the environmental predictors with GAMs using a normal distribution and four degrees of freedom. The fit of the SR and functional models were calculated with D^2 and their predictive power by computing a Spearman rank correlation between the observed and predicted indices values for the evaluation dataset.

S-SDM constraints, species selection and assemblage evaluation

The following analyses were performed on the evaluation dataset. In addition to the unconstrained S-SDM predictions, we tested five different ways of constraining S-SDMs, based on two main approaches:

1) Limiting SR so that it fits the corresponding MEM prediction; we used two main rules to select the species in each plot up to the constrained number: a random selection of species; and a selection of the species in decreasing order of their SDM predicted probability of presence.

2) Using the predicted values of the 5^{th} and 95^{th} percentiles of H, SLA and both, as criterion to discard those species considered as functionally outside of the assemblage.

This yielded six different S-SDM predictions: (i) unconstrained stacked species distribution models (S-SDM.), (ii) SR random (SRrand), (iii) SR ordered probabilities

(SRprob), (iv) min-max of H (H), (v) min-max of SLA (SLA), and (vi) min-max of H and SLA (HSLA).

The quality of the predicted assemblage was evaluated with several metrics based on a confusion matrix (Table 1) into which the total number of species used in the analysis (species pool: *SP*) are classified into: *TP:* the species observed as well as predicted as present (true positive), *FN:* the species observed as present but predicted as absent (false negative; omission error), *FP:* the species observed as absent but predicted as present (false positive; commission error) and *TN:* the species both observed and predicted as absent (true negative).

Table 1: Confusion matrix used to compute the assemblage evaluation metrics

		observed	
		0	**1**
predicted	**0**	*TN*	*FN*
	1	*FP*	*TP*

We computed the species richness error (predicted SR – observed SR), the assemblage prediction success (a), and the Sørensen index (b), also known as Bray-Curtis similarity.

$$(a)\ Prediction\ success = \frac{TP + TN}{SP}$$

$$(b)\ Sørensen\ index = \frac{2TP}{2TP+FN+FP}$$

RESULTS

Evaluation results for the individual SDMs and MEMs are presented in supplementary information. As results for community predictions using TSS were similar to those using AUC, they are presented in supplementary information (Figures S1 and S2).

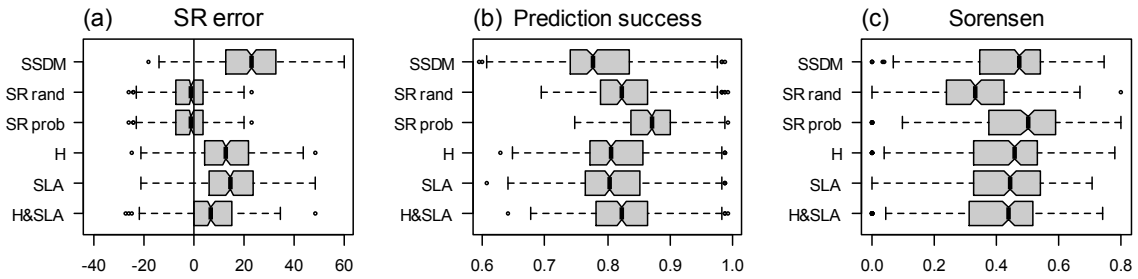


Figure 1: Boxplots summarizing assemblage prediction evaluation, (a) species richness error, (b) prediction success and (c) Sorensen index; for S-SDM without limitation and S-SDM constrained by SR with a random choice of kept species (SR rand), S-SDM constrained by SR keeping only the most probable species (SR prob), S-SDM constrained by predicted minimum and maximum vegetative height of the community (H), specific leaf area of the community (SLA) and both (H&SLA).

Consistently with what was found in Dubuis et al. (2011), the S-SDM overpredicted species richness (SR) in each plot (Fig. 1a). All filtering types contributed on average to reduce SR overprediction, and to increase prediction success (Fig. 1b). As to composition predictions, the Sørensen index (Fig. 1c) decreased in all but one case, when using predicted SR as a limit and when selecting species in decreasing order of predicted probability of presence (SRprob). In the latter case, the Sørensen index was significantly higher than the one of SSDM (Wilcoxon signed rank test, p-value=0.004). On average, this SRprob approach was less affected by errors of commission (false positive; table x) than other approaches (Fig. 2).

Using SR as a limit but choosing species randomly among those predicted yielded the worst assemblage prediction. Predicted functional traits were less constraining and did not allow for a complete reduction of the SR over-prediction. Their use allowed increasing prediction success, but at the cost of decreasing slightly the Sorensen index.

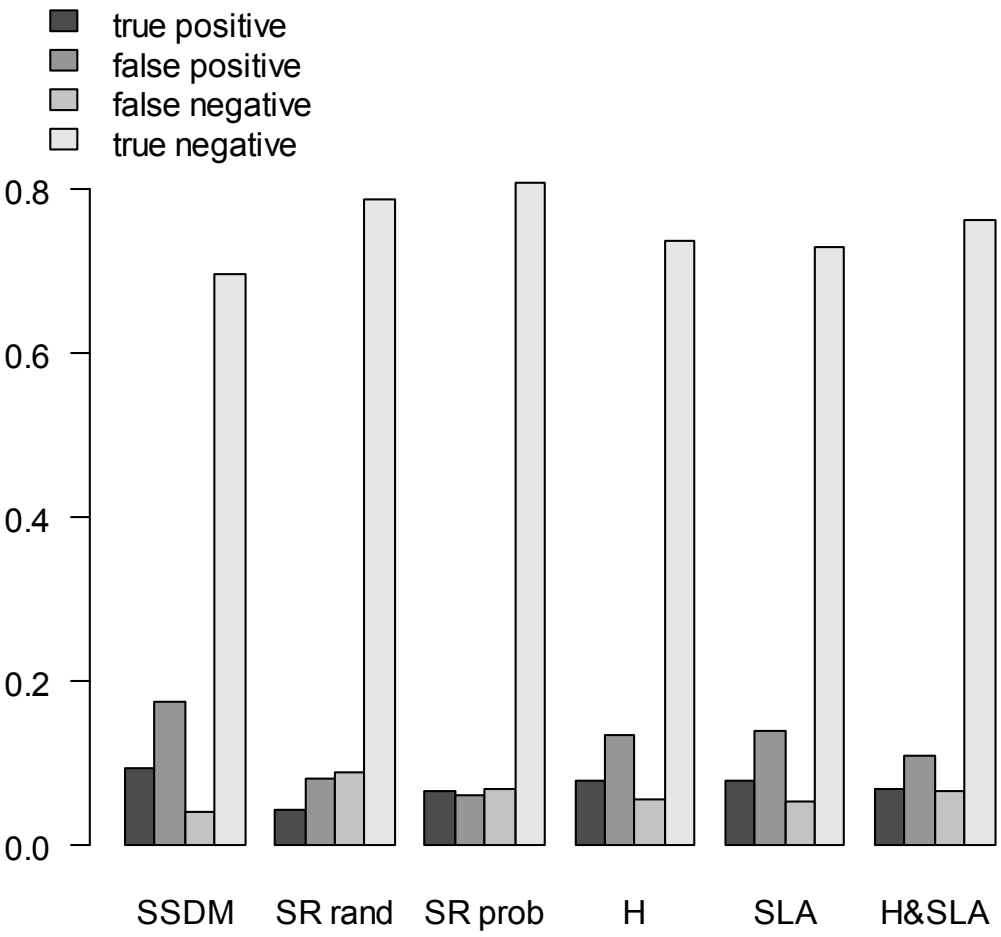


Figure 2: Barplot showing the proportion (mean among all plots) of true and false positive, as well as true and false negative for S-SDM for each of the five limitations.

DISCUSSION

In our study, applying macro-ecological constraints on S-SDMs yielded informative results regarding the improvement of the overall quality of species assemblage predictions. Constraining S-SDM with MEM output reduced the over-prediction of SR in all cases, providing support to this important dimension of the SESAM framework. Using symmetric evaluation measures weighting errors of commission and omission equally, it also improved prediction success in all cases. However, among all approaches tested, only one allowed improving community predictions based on composition similarity: using the macroecological model for SR as limit for the number of species occurring in the assemblage and retaining the species with their highest predicted probability of presence. Functional traits models outputs were not as efficient to constrain the prediction, probably because we considered only two traits. Using others or more functional traits could increase the ecological significance and efficiency of the predictions. Species assemblage modelling approaches based on traits use more comprehensive trait datasets (Shipley et al., 2011). Seeds and roots traits could be good candidates to complete the functional constraints applied to the model (Dubuis et al., in review). A potential limitation to the use of particular traits is that they must be predicted reliably in space by MEM (e.g., Dubuis et al., in review), which is not always possible.

Overall, and even after limiting the SR overprediction, predicted assemblage composition was still significantly distinct from the observed one, consistent with the results by Aranda & Lobo (2011) and Pottier et al. (2012). S-SDMs as developed with the data at hand were not able to predict species assemblages with an average accuracy greater than a Sorensen of 0.5. Even if the SDM used to build the assemblage prediction have reasonably good independent evaluation values, they nevertheless contain errors, which can furthermore be unevenly distributed across the environment (Pottier et al. in press). While stacking SDMs, these errors accumulate, accordingly degrading the assemblage prediction.

In this regard, the SESAM framework provides some promising perspectives, by allowing decreasing the rate of commission error.

Nevertheless, one particular problem remains that should be reduced before continuing research in this direction: the omission error rate in SDMs and in the SSDM. Whereas S-SDMs were shown to overpredict the number of species, individual species can nevertheless also be omitted. Although our current implementation of SESAM

framework can reduce errors of commission, it cannot prevent omission errors, i.e., adding unpredicted species to an assemblage.

A likely source of omission errors in our case may come from limitations related to the environmental predictors used to build the SDMs. Available predictors can include some level of errors (e.g. from interpolation) and other important predictors may likely be missing in the underlying SDMs. As a result, species' realized niches are likely incompletely described and suitable situations for a species are not captured in the model. Two recent papers have shown similar problems of assemblage predictions in the case of plants and butterflies respectively (Pellissier et al., 2011; Pottier et al., 2012). In both cases, assemblage prediction sensitivity was lower at the highest elevations, supposedly due to the less stable environmental conditions there and to missing substrate predictors (e.g. rock type, soil depth). This problem could be solved at least partially by using more proximal and spatially accurate environmental predictors. Regarding our study area, snow cover and geomorphology (Randin et al., 2009), soil moisture and soil temperature (Le Roux et al., 2013) as well as edaphic conditions (Dubuis et al., 2012) could help ameliorate the SDMs and hence SSDMs.

The fourth component of the SESAM framework, i.e. including biotic rules to choose between species, has not been tested in this study. Applying ecological assembly rules for selecting species co-occurring in an assemblage may additionally help decrease the omission rate. However, identifying and quantifying ecological assembly rules that can be applied generally appears difficult with the current state of knowledge (Götzenberger et al., 2011, Wisz et al. 2012).

We have shown how S-SDM predictions can be improved by applying an MEM constraint to species richness, and then selecting species by decreasing order of predicted probability (until the MEM richness is reached). Yet, SDMs have errors, unevenly distributed in geographic and environmental space, that cumulate into S-SDMs. Furthermore, niche-based processes are only one type of processes contributing to assemblage formation, together with stochastic demographic, dispersal and disturbance processes (Leibold & McPeek, 2006; Adler et al., 2007). Therefore, assemblage composition will always entail some level of prediction errors. What will prove useful in future studies will be to better understand and differentiate the different sources of errors in SSDMs, as e.g. due to limitations in SDMs or to non-niche stochastic processes. By extension, it would be useful to assess how errors propagate from individual SDMs to S-SDMs, and what value of the Sørensen index (or other evaluation metric of communities) would qualify a fair assemblage prediction. This will

help evaluate how the SESAM framework may contribute to a better understanding of community assembly in space and time.

ACKNOWKEDGMENTS

We are grateful to the numerous people who contributed to the data collection and to Glenn Litsios for insightful discussions and comments on the manuscript. This study was supported by the Swiss National Science Foundation (grant Nr. 31003A-125145 to A. Guisan) and by the FP6 Ecochange project of the European Commission (grant GOCE-CT-2007–036866).

REFERENCES

Adler P.B., Hillerislambers J., & Levine J.M. (2007) A niche for neutrality. *Ecology Letters*, **10**, 95–104.

Albouy C., Guilhaumon F., Araújo M.B., Mouillot D., & Leprieur F. (2012) Combining projected changes in species richness and composition reveals climate change impacts on coastal Mediterranean fish assemblages. *Global Change Biology*, **18**, 2995–3003.

Algar A.C., Kharouba H.M., Young E.R., & Kerr J.T. (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography*, **32**, 22–33.

Allouche O., Tsoar A., & 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.

Aranda S.C. & Lobo J.M. (2011) How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. *Ecography*, **34**, 31–38.

Baselga A. & Araújo M.B. (2009) Individualistic vs community modelling of species distributions under climate change. *Ecography*, **32**, 55–65.

Baselga A. & Araújo M.B. (2010) Do community-level models describe community variation effectively? *Journal of Biogeography*, no–no.

Dubuis A., Giovanettina S., Pellissier L., Pottier J., Vittoz P., & Guisan A. (2012) Soil pH and nitrogen content data improve the predictions of plant species distributions and community composition in a mountainous landscape. *Journal of Vegetation Science*, n/a–n/a.

Dubuis A., Pottier J., Rion V., Pellissier L., Theurillat J.-P., & Guisan A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, no–no.

Dubuis A., Rossier L., Pottier J., Pellissier L., & Guisan A. Predicting current and future community patterns of plant functional traits. *Ecography*, .

Ferrier S. & Guisan A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.

Francis A.P. & Currie D.J. (2003) A globally consistent richness-climate relationship for angiosperms. *The American naturalist*, **161**, 523–36.

Guisan A. & Rahbek C. (2011) SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, no–no.

Götzenberger L., De Bello F., Anne Bråthen K., Davison J., Dubuis A., Guisan A., Lepš J., Lindborg R., Moora M., Pärtel M., Pellissier L., Pottier J., Vittoz P., Zobel K., & Zobel M. (2011) Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biological reviews of the Cambridge Philosophical Society*, .

Leibold M.A. & McPeek M.A. (2006) Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, **87**, 1399–410.

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

Mateo R.G., Felicísimo Á.M., Pottier J., Guisan A., & Muñoz J. (2012) Do stacked species distribution models reflect altitudinal diversity patterns? *PloS one*, **7**, e32586.

Mokany K., Harwood T.D., Overton J.M., Barker G.M., & Ferrier S. (2011) Combining α- and β-diversity models to fill gaps in our knowledge of biodiversity. *Ecology letters*, **14**, 1043–51.

Mokany K., Harwood T.D., Williams K.J., & Ferrier S. (2012) Dynamic macroecology and the future for biodiversity. *Global Change Biology*, **18**, 3149–3159.

Moser D., Dullinger S., Englisch T., Niklfeld H., Plutzar C., Sauberer N., Zechmeister H.G., & Grabherr G. (2005) Environmental determinants of vascular plant species richness in the Austrian Alps. *Journal of Biogeography*, **32**, 1117–1127.

Pellissier L., Pradervand J.-N., Pottier J., Dubuis A., Maiorano L., & Guisan A. (2011) Climate-based empirical models show biased predictions of butterfly communities along environmental gradients. *Ecography*, no–no.

Pottier J., Dubuis A., Pellissier L., Maiorano L., Rossier L., Randin C.F., Vittoz P., & Guisan A. (2012) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, **In press**, n/a–n/a.

Randin C.F., Vuissoz G., Liston G.E., Vittoz P., & Guisan A. (2009) Introduction of Snow and Geomorphic Disturbance Variables into Predictive Models of Alpine Plant Distribution in the Western Swiss Alps. *Arctic, Antarctic, and Alpine Research*, **41**, 347–361.

Le Roux P.C., Lenoir J., Pellissier L., Wisz M.S., & Luoto M. (2013) Horizontal, but not vertical, biotic interactions affect fine-scale plant distribution patterns in a low energy system. *Ecology*, .

Shipley B., Laughlin D.C., Sonnier G., & Ottinowski R. (2011) A strong test of a maximum entropy model of trait-based community assembly. *Ecology*, **92**, 507–17.

Shipley B., Vile D., & Garnier E. (2006) From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science*, **314**, 812–4.

Sonnier G., Shipley B., & Navas M. (2010a) Plant traits , species pools and the prediction of relative abundance in plant communities : a maximum entropy approach. *Journal of Vegetation Science*, 318–331.

Sonnier G., Shipley B., & Navas M.L. (2010b) Quantifying relationships between traits and explicitly measured gradients of stress and disturbance in early successional plant communities. *Journal of Vegetation Science*, **21**, 318–331.

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

SUPPLEMENTARY INFORMATIONS

Table S1: Results of evaluation for macroecological models for functional traits quantiles and species richness.

	D2	Correlation
5th logH	0.627	0.794
5th logSLA	0.687	0.833
95th logH	0.736	0.859
95th logSLA	0.503	0.716
Species richness	0.431	0.542

Table S2: Summary of the SDMs evaluation results.

	GAM	GBM	GLM
AUC mean	0.801	0.790	0.798
AUC stdev	0.077	0.076	0.077
TSS mean	0.538	0.514	0.532
TSS stdev	0.139	0.138	0.143

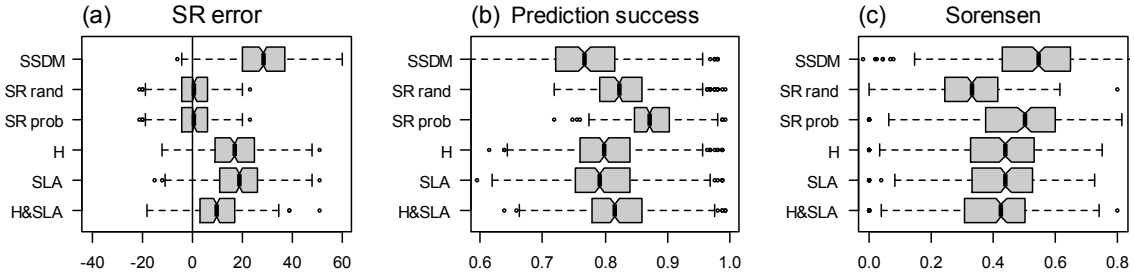


Figure S1: Boxplots summarizing assemblage prediction evaluation for S-SDMs based on TSS, (a) species richness error, (b) prediction success and (c) Sorensen index; for S-SDM without limitation and S-SDM constrained by SR with a random choice of kept species (SR rand), S-SDM constrained by SR keeping only the most probable species (SR prob), S-SDM constrained by predicted minimum and maximum vegetative height of the community (H), specific leaf area of the community (SLA) and both (H&SLA).

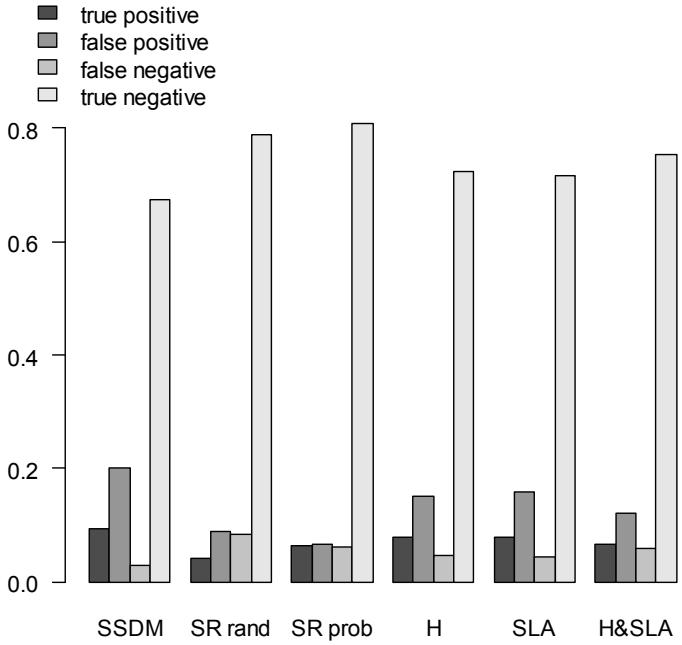
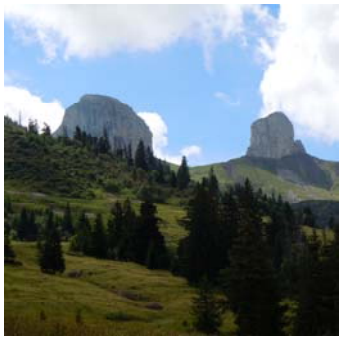


Figure S2: Barplot showing the proportion (mean among all plots) of true and false positive, as well as true and false negative for S-SDM based on TSS for each of the five limitations.

SYNTHESIS



In this thesis, I developed and tested new environmental variables as predictors in SDMs. I developed and tested MEMs for species richness and functional assemblage properties. Finally, I combined S-SDMs and MEMs to conduct a first test of the SESAM framework that aims at obtaining more precise predictions of plant assemblages composition.

Developing and testing new environmental predictors

Throughout this thesis I have highlighted several important aspects for the further development of SDMs and assemblage modelling. First of all, while it has been known for a long time (Guisan & Zimmermann, 2000; Austin, 2007) and shown in numerous previous papers (Randin et al., 2009a, 2009b), the need of environmental predictors at sufficient precision and having a biological meaning remains a current and partly unsolved matter. In chapter 1, we underlined the importance of introducing edaphic predictors more systematically in SDMs for herbaceous species. Edaphic variables allowed more precise habitat discrimination for species showing preference for acidic soils. This corroborates results from studies done on some shrubs, trees, and forest communities (Coudun et al., 2006; Coudun & Gégout, 2007; Marage et al., 2009). A recent study also evidenced the importance of taking edaphic variables into account when predicting the impact of climate change on a species distribution such as not overestimate future distributions (Bertrand et al., 2012).

In chapter 4, the same edaphic variables were used to predict spatial patterns of functional traits. They were of particular importance for modelling the traits related to resource acquisition and management such as leaf dry matter content or leaf nitrogen content. Models for those traits could not be projected in space or in the future due to the lack of edaphic predictors covering the whole study area. Unfortunately such predictors that could help ameliorate SDMs by complementing more classical topo-climatic predictors are often not available. Soil pH, the most important edaphic variable for the species of our dataset, for example does not exist as a spatially explicit map for our study area. Several studies have attempted to model pH data to create maps (Shi et al., 2009; Castrignanò et al., 2011). We carried out tests to predict pH in our study area, but the results were not sufficient. As our study area presents a high level of topographic heterogeneity the use of kriging techniques for interpolation proved inefficient. We also tested regression based techniques (McKenzie & Ryan, 1999) that predict pH as a function of environmental variable (bedrock, elevation, curvature) without success. The obtained pH map has proved too smooth to be used as an efficient predictor in SDMs (Boheler et al. unpublished). We concluded that, soil depth

data would be needed to better account for the influence of bed rock or loess accumulation on local pH. Consequently, more work would be needed to produce good quality edaphic maps for our study area.

In chapter 2 (Pradervand et al. in prep) we demonstrated that the use of fine resolution topographic predictors did not allow a constant improvement of predictive power of SDMs. On the one hand, this demonstrates that the predictors used in previous studies were valid. On the other hand, it indicates that more work is needed to obtain topographic and climatic predictors better reflecting conditions experienced by plants. The use of micro-loggers to register temperature data at the soil level seems to be promising (Scherrer & Körner, 2011; Scherrer et al., 2011) even if constraining due to the number of necessary loggers to obtain enough precision. Direct measurement of the predictors in the field allow an unmatched precision (Le Roux et al., 2013), but this is not possible for large study area. A compromise must be found to maximize the quality of the environmental variables and surface on which they can be interpolated with precision. Predictors resulting from remote-sensing data could also represent a promising way to obtain high resolution climatic data (Pouteau et al., 2012).

Biotic predictors are another class of predictors that have increasingly proven to be important for SDMs. Presences of one or several dominant species have been used successfully to account for biotic interactions limiting or promoting the presence of the modelled species (Pellissier et al. 2010, annex 1; Le Roux et al. 2013). The use of dominant species distribution as predictor requires advanced ecological knowledge of the relations between species in the study area. However, such a strategy can only be implemented in a biological system where sufficiently dominant species occurs and can be used as predictor which is not the case in our study area. Alternatives to this limitation could be to test a large set of wide spread species as predictor (Le Roux et al., 2012) or to use more general co-occurrence-based indices (Boulangeat et al., 2012). The importance of biotic interaction in explaining species distribution can also be illustrated with the case of *Leontopodium alpinum* (Ischer et al in review, annex 5) that tracks the climatic and edaphic conditions allowing it to avoid too productive habitats and consequently too dense competing vegetation. Taking biotic predictors into account in this case could be a way to obtain a more precise prediction of the species distribution. Comprehensive reviews of the use of biotic interaction in SDMs have been proposed by Wisz et al. (2012) and Kissling et al. (2011). Potential drawbacks associated with the use of biotic predictors are that the latter are likely to be

influenced by other factors, such as land-use that has important effects in our study area and can vary in space and time.

The results obtained for species assemblage modelling also gave evidence that pursuing development of new spatially-explicit and biologically relevant predictors is a task of high importance. This is discussed in more details below.

Developing macro-ecological models

The second achievement of this thesis has been to develop and test macro-ecological models of plant assemblage attributes. In chapter 3, I investigated the possibility of predicting species richness from environmental predictors directly using predictive models with a Poisson distribution. I then compared the obtained results with the observed species richness and the one resulting from a stacking of individual SDMs. I showed that similarly to previous studies (Algar et al., 2009; Newbold et al., 2009; Pineda & Lobo, 2009), species richness predicted with S-SDM was higher than the observed richness. This was not the case for species richness predicted with MEM, which predicted a realistic number of species and also better reproduced the observed hump-shaped species richness pattern along the elevation gradient. These results indicate that the predictors used in our study reflect well the factors controlling species richness. The resulting projection could thus, at least theoretically, be used to limit S-SDM over-prediction as proposed in the SESAM framework (Guisan & Rahbek, 2011).

In chapter 4, I have explored the possibility of using MEM to predict functional characteristics of plant communities through functional indices. This study showed that indices summarizing the trait values of a species assemblage, like community weighted mean or extreme values of traits could be predicted with a good accuracy. On the contrary, diversity indices that picture the way functional traits are organized inside the species assemblage are much more difficult to predict accurately with climatic, topographic and edaphic predictors. They are likely dependant of other environmental factors like disturbance, stress and land-use that were not considered in our study. Chance is also likely playing a role. Despite that, weighted mean and extreme values of traits in a community show promising possibility. Extreme traits values that can be predicted with good accuracy reflect environmental filtering effects, allowing to discriminate species that can occur in given environmental condition from those that cannot. This could be of particular interest in the context of the SESAM framework.

Assemblage composition prediction with S-SDMs

A third achievement of this thesis has been the test of the ability of S-SDMs to predict the composition of species assemblages. We showed that, when stacking models that alone have good predictive accuracy (measured with conventional evaluation metrics like AUC) the resulting assemblage predictions contained a considerable proportion of error, either for plants (Pottier et al. 2012, annex 2; Dubuis et al. in prep; chapter 5) or for butterflies (Pellissier et al. 2011, annex 3). As seen in chapter 3, S-SDMs tend to over-predict species richness, and the analyses carried out in chapter 5 showed that the proportion of omission errors in assemblage prediction is also important. These errors come from SDMs and can be explained partly by the lack of more suitable environmental predictors or from inaccuracies in those used (Pottier et al., 2013). For example, models for butterfly assemblage would require the addition of predictors reflecting the vegetation diversity (Pellissier et al. 2011, annex 3). Plant models would benefit from more proximal topo-climatic predictors like snow cover or geomorphology at high elevation and from land-use at low elevation, as such factors can have more influence on vegetation than the climate in some places (Pottier et al. 2012, annex 2). Prediction errors in SDMs can also be explained by sampling errors during data collection, occurring species can be overlooked in a favourable habitat and cause over-prediction during modelling (Chen et al., 2013). Aranda & Lobo (2011), for example, explained the unsatisfactory quality of their species assemblage prediction with the bad quality of data used to build the SDM. Finally, population dynamics and the related stochastic processes that are not niche-based and contradict pseudo-equilibrium assumption of SDMs also play a role in the error accumulation. This problem may be the most difficult to solve (Guisan & Thuiller, 2005).

SESAM framework implementation

In chapter 5, I proposed a first partial test of the SESAM framework (Guisan & Rahbek, 2011) that aims to improve results from S-SDM. We used a model for species richness and models for trait extremes values, such as those developed in chapter 3 and 4, to limit the number of species predicted in each plots. SR limitation gave the best result, in term of species number and assemblage composition. Limitations based on functional traits were not restrictive enough. Clearly, more traits would be necessary to obtain a more realistic species assemblage. It would be interesting to consider other assemblage properties as limitations such as for example biomass, or cover of species.

Contrarily to what was suggested in the original SESAM paper we did not use ecological assembly rules to choose between species predicted as present. Ecological assembly rules are difficult to detect and generalize (Götzenberger et al. 2011, annex 4). Indeed, we used the simple rule of selecting the species based on their probability of presence predicted by the model. This method appears to be efficient, but more research would be needed at this stage. Using extreme trait values can be seen as a kind of assembly rules as it reflects environmental filtering on species. SESAM framework seems promising, as limiting the number of species predicted by S-SDM with a species richness model improved the prediction success as well as the similarity of the prediction with observed data. But if SESAM framework allows to decrease commission errors, it cannot solve omission errors which proportionally is important, as also shown in other studies (Pineda & Lobo, 2009; Aranda & Lobo, 2011). Consequently, before going further in SESAM implementation, SDMs should be improved as much as possible.

Perspectives

SDMs that are intended for S-SDM building must be formulated carefully as the error they contain will accumulate when stacked. As we have seen previously, the addition of more proximal predictors to SDMs to better estimate the realized niche of the species is a first necessary measure. Building SDMs not only with presences and absences but with abundances data could provide a more precise view of the species distribution as well as avoid some overestimation of the areas predicted as suitable (Van Couwenberghe et al., 2013). The quality of the species data used to build the model should also be considered carefully. But unfortunately, exhaustive sampling is not possible and errors are likely to occur in every existing data-set (Vittoz & Guisan, 2007). Likewise, estimating on the field the status of a population (source or sink) looks difficult as part of vegetation sampling, without setting up experiments. Regarding processes that are not niche based, numerous approaches have been proposed to integrate mechanistic models of dispersal, landscape or population dynamics to SDMs, allowing improved species distribution prediction (see Franklin 2010 for a review; Pagel & Schurr 2012; Swab et al. 2012 for recent applications). While having shown to be efficient, these approaches remain data demanding, needing known life-history parameters for each modelled species (Franklin, 2010). Such approaches may thus be difficult to apply with the aim of predicting species assemblages for an important species pool. Moreover, as these methods are intended to analyse species distribution in the future, they need to simulate the species dynamic in time, and thus will not be

readily useful to ameliorate the prediction of many species distributions in space at the present time. Recently, Mokany et al. presented a semi-mechanistic approach to predict species assemblage composition in space and time (Mokany & Ferrier, 2011; Mokany et al., 2011). Their framework merge modelling of community level attributes (species richness and beta diversity), with occurrences data and a dynamic optimization algorithm to predict the composition of species assemblage in space. They showed that their approach could also predict species assemblages under climate change scenarios (Mokany et al., 2012). This method seems promising and needs to be tested on other data sets.

Despite their drawbacks, S-SDMs and by extension SESAM remain interesting approaches to predict species assemblage allowing precise prediction and relative ease of use. Finally, one important question that remains to be assessed is what proportion of a species assemblage cannot be predicted (due to stochastic processes for example). Answering that question, through simulations for example, would assess what similarity of composition could be expected when SDM account for all empirical factors, and thus tell when an assemblage prediction can be considered as reliable.

REFERENCES

Algar A.C., Kharouba H.M., Young E.R., & Kerr J.T. (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography*, **32**, 22–33.

Aranda S.C. & Lobo J.M. (2011) How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. *Ecography*, **34**, 31–38.

Austin M. (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–19.

Bertrand R., Perez V., & Gégout J.-C. (2012) Disregarding the edaphic dimension in species distribution models leads to the omission of crucial spatial information under climate change: the case of Quercus pubescens in France. *Global Change Biology*, **18**, 2648–2660.

Boulangeat I., Gravel D., & Thuiller W. (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters*, **15**, 584–93.

Castrignanò A., Buttafuoco G., & Comolli R. (2011) Using Digital Elevation Model to Improve Soil pH Prediction in an Alpine Doline. *Pedosphere*, **21**, 259–270.

Chen G., Kéry M., Plattner M., Ma K., & Gardner B. (2013) Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology*, **101**, 183–191.

Coudun C. & Gégout J.-C. (2007) Quantitative prediction of the distribution and abundance of Vaccinium myrtillus with climatic and edaphic factors. *Journal of Vegetation Science*, **18**, 517–524.

Coudun C., Gégout J.-C., Piedallu C., Rameau J.-C.C., & Gegout J.C. (2006) Soil nutritional factors improve models of plant species distribution: an illustration with Acer campestre (L.) in France. *Journal of Biogeography*, **33**, 1750–1763.

Van Couwenberghe R., Collet C., Pierrat J.-C., Verheyen K., & Gégout J.-C. (2013) Can species distribution models be used to describe plant abundance patterns? *Ecography*, in press.

Franklin J. (2010) Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions*, **16**, 321–330.

Guisan A. & Rahbek C. (2011) SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.

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.

Götzenberger L., De Bello F., Anne Bråthen K., Davison J., Dubuis A., Guisan A., Lepš J., Lindborg R., Moora M., Pärtel M., Pellissier L., Pottier J., Vittoz P., Zobel K., & Zobel M. (2012) Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biological Reviews*, **87**, 111–127.

Kissling W.D., Dormann C.F., Groeneveld J., Hickler T., Kühn I., McInerny G.J., Montoya J.M., Römermann C., Schiffers K., Schurr F.M., Singer A., Svenning J.-C., Zimmermann N.E., & O'Hara R.B. (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Marage D., Gégout J., & Gegout J.C. (2009) Importance of soil nutrients in the distribution of forest communities on a large geographical scale. *Global Ecology and Biogeography*, **18**, 88–97.

McKenzie N.J. & Ryan P.J. (1999) Spatial prediction of soil properties using environmental correlation. *Geoderma*, **89**, 67–94.

Mokany K. & Ferrier S. (2011) Predicting impacts of climate change on biodiversity: a role for semi-mechanistic community-level modelling. *Diversity and Distributions*, **17**, 374–380.

Mokany K., Harwood T.D., Overton J.M., Barker G.M., & Ferrier S. (2011) Combining α - and β -diversity models to fill gaps in our knowledge of biodiversity. *Ecology Letters*, **14**, 1043–51.

Mokany K., Harwood T.D., Williams K.J., & Ferrier S. (2012) Dynamic macroecology and the future for biodiversity. *Global Change Biology*, **18**, 3149–3159.

Newbold T., Gilbert F., Zalat S., El-Gabbas A., & Reader T. (2009) Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography*, **36**, 2085–2095.

Pagel J. & Schurr F.M. (2012) Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, **21**, 293–304.

Pellissier L., Bråthen K.A., Pottier J., Randin C.F., Vittoz P., Dubuis A., Yoccoz N.G., Alm T., Zimmermann N.E., Guisan A., & Brathen K.A. (2010) Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, **33**, 1004–1014.

Pellissier L., Pradervand J.-N., Pottier J., Dubuis A., Maiorano L., & Guisan A. (2012) Climate-based empirical models show biased predictions of butterfly communities along environmental gradients. *Ecography*, **35**, 684–692.

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

Pottier J., Dubuis A., Pellissier L., Maiorano L., Rossier L., Randin C.F., Vittoz P., & Guisan A. (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, **22**, 52–63.

Pouteau R., Meyer J.-Y., Taputuarai R., & Stoll B. (2012) Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecological Informatics*, **9**, 37–46.

Randin C.F., Jaccard H., Vittoz P., Yoccoz N.G., & Guisan A. (2009a) Land use improves spatial predictions of mountain plant abundance but not presence-absence. *Journal of Vegetation Science*, **44**, 996–1008.

Randin C.F., Vuissoz G., Liston G.E., Vittoz P., & Guisan A. (2009b) Introduction of Snow and Geomorphic Disturbance Variables into Predictive Models of Alpine Plant Distribution in the Western Swiss Alps. *Arctic, Antarctic, and Alpine Research*, **41**, 347–361.

Le Roux P.C., Lenoir J., Pellissier L., Wisz M.S., & Luoto M. (2013) Horizontal, but not vertical, biotic interactions affect fine-scale plant distribution patterns in a low energy system. *Ecology*, in press.

Le Roux P.C., Virtanen R., Heikkinen R.K., & Luoto M. (2012) Biotic interactions affect the elevational ranges of high-latitude plant species. *Ecography*, **35**, 1048–1056.

Scherrer D. & Körner C. (2011) Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *Journal of Biogeography*, **38**, 406–416.

Scherrer D., Schmid S., & Körner C. (2011) Elevational species shifts in a warmer climate are overestimated when based on weather station data. *International journal of biometeorology*, **55**, 645–54.

Shi W.J., Liu J.Y., Du Z.P., Song Y.J., Chen C.F., & Yue T.X. (2009) Surface modelling of soil pH. *Geoderma*, **150**, 113–119.

Swab R.M., Regan H.M., Keith D. a., Regan T.J., & Ooi M.K.J. (2012) Niche models tell half the story: spatial context and life-history traits influence species responses to global change. *Journal of Biogeography*, **39**, 1266–1277.

Vittoz P. & Guisan A. (2007) How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. *Journal of Vegetation Science*, **18**, 413–422.

Wisz M.S., Pottier J., Kissling W.D., Pellissier L., Lenoir J., Damgaard C.F., Dormann C.F., Forchhammer M.C., Grytnes J.-A., Guisan A., Heikkinen R.K., Høye T.T., Kühn I., Luoto M., Maiorano L., Nilsson M.-C., Normand S., Ockinger E., Schmidt N.M., Termansen M., Timmermann A., Wardle D. a, Aastrup P., & Svenning J.-C. (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological reviews of the Cambridge Philosophical Society*, **88**, 15–30.

ANNEX 1

Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants

Ecography 33: 1004–1014, 2010
doi: 10.1111/j.1600-0587.2010.06386.x

Subject Editors: Jens-Christian Svenning and Peter Pearmen. Accepted 7 September 2010

Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants

Loïc Pellissier, Kari Anne Bråthen, Julien Pottier, Christophe F. Randin, Pascal Vittoz, Anne Dubuis, Nigel G. Yoccoz, Torbjørn Alm, Niklaus E. Zimmermann and Antoine Guisan

L. Pellissier (loic.pellissier@unil.ch), J. Pottier, C. F. Randin, P. Vittoz, A. Dubuis and A. Guisan, Univ. of Lausanne, Dept of Ecology and Evolution (DEE), Faculty of Biology and Medicine, CH-1015 Lausanne, Switzerland. (Present address of P. V. and A. G.: Univ. of Lausanne, Faculty of Geosciences and Environment, CH-1015 Lausanne, Switzerland.) – K. A. Bråthen and N. G. Yoccoz, Dept of Arctic and Marine Biology, Univ. of Tromsø, NO-9037 Tromsø, Norway. – T. Alm, Tromsø Museum, Univ. of Tromsø, NO-9037 Tromsø, Norway. – N. E. Zimmermann, Swiss Federal Research Inst. WSL, Land Use Dynamics Unit, CH-8903 Birmensdorf, Switzerland.

Abiotic factors are considered strong drivers of species distribution and assemblages. Yet these spatial patterns are also influenced by biotic interactions. Accounting for competitors or facilitators may improve both the fit and the predictive power of species distribution models (SDMs). We investigated the influence of a dominant species, *Empetrum nigrum* ssp. *hermaphroditum*, on the distribution of 34 subordinate species in the tundra of northern Norway. We related SDM parameters of those subordinate species to their functional traits and their co-occurrence patterns with *E. hermaphroditum* across three spatial scales. By combining both approaches, we sought to understand whether these species may be limited by competitive interactions and/or benefit from habitat conditions created by the dominant species. The model fit and predictive power increased for most species when the frequency of occurrence of *E. hermaphroditum* was included in the SDMs as a predictor. The largest increase was found for species that 1) co-occur most of the time with *E. hermaphroditum*, both at large (i.e. 750 m) and small spatial scale (i.e. 2 m) or co-occur with *E. hermaphroditum* at large scale but not at small scale and 2) have particularly low or high leaf dry matter content (LDMC). Species that do not co-occur with *E. hermaphroditum* at the smallest scale are generally palatable herbaceous species with low LDMC, thus showing a weak ability to tolerate resource depletion that is directly or indirectly induced by *E. hermaphroditum*. Species with high LDMC, showing a better aptitude to face resource depletion and grazing, are often found in the proximity of *E. hermaphroditum*. Our results are consistent with previous findings that both competition and facilitation structure plant distribution and assemblages in the Arctic tundra. The functional and co-occurrence approaches used were complementary and provided a deeper understanding of the observed patterns by refinement of the pool of potential direct and indirect ecological effects of *E. hermaphroditum* on the distribution of subordinate species. Our correlative study would benefit being complemented by experimental approaches.

Understanding the relative roles of climate and species interactions in shaping species ranges is among the oldest challenges in ecology and has now become a prerequisite for predicting species' responses to global change (Adler and HilleRisLambers 2008). A large number of studies have focused on abiotic drivers of species distributions, such as climatic or land-use factors (Randin et al. 2009; see Guisan and Zimmermann 2000). However, it is also recognised that the observed patterns of species occurrences are strongly influenced by source-sink dynamics, dispersal limitation and biotic interactions (Pulliam 2000, Lortie et al. 2004, Gotelli et al. 2010) which should be taken into consideration in species distribution studies (Guisan and Thuiller 2005). In particular, predation, competition or facilitation may restrict or inflate species ranges (Pulliam 2000). Therefore, species interactions shaping community assemblages should also be accounted for.

Recent analyses using species distribution models (SDMs) have shown that the inclusion of additional predictors that account for known competitors or facilitators may increase both our understanding of species range limits and the predictive power of models (Leathwick and Austin 2001, Anderson et al. 2002, Heikkinen et al. 2007, Meier et al. 2010). However, positive or negative relationships between species inferred from such empirical models only provide limited information. Deeper ecological insights regarding species relationships are likely to come from other approaches developed in community ecology. For instance, the analysis of species co-occurrence patterns has been used to assess the importance of biotic interactions

and to evaluate whether biotic assembly rules determine the structure of natural communities (i.e. if species co-occur less or more frequently than would be expected by chance alone; Weiher and Keddy 1999, Gotelli and McCabe 2002). However, the detection of co-occurrence patterns and the ecological processes behind them is scale dependent (Weiher and Keddy 1999), and thus, such analyses may be refined by measuring and comparing co-occurrences at different scales. Additionally, the study of co-occurrence alone might be meaningless when considering species altogether, irrespective of their ecological characteristics. That is, co-occurrence analyses and SDMs provide correlative inferences that do not necessarily relate to the causal effects of competition or facilitation on species ranges.

One proposed alternative is to use functionally relevant plant traits that provide means of analysing biotic interactions from trait similarities or differences between species (Lavorel et al. 1997, Kraft et al. 2008). Here we assume that trait patterns can provide useful insights about the role of interactions in shaping actual spatial associations between species (Keddy et al. 1998, Wardle et al. 1998); that is, the functional approach might provide complementary insights into the statistical relationships between species found in SDMs and/or within co-occurrence analyses. While this approach is also correlative in nature, we argue that combining SDMs with approaches inherited from community ecology can deepen our understanding of the relationships that shape species distributions.

Although competition was long considered to be the main type of biotic interaction driving species assemblages (Tilman 1994), during the last decade, numerous experimental studies demonstrated that positive interactions may also play an important role in these processes, particularly in severe environments (Choler et al. 2001, Callaway et al. 2002), which is also reflected in the stress-gradient hypothesis (Bertness and Callaway 1994). This hypothesis predicts that the frequency of facilitative and competitive interactions will vary inversely across abiotic stress gradients with facilitation being more common in conditions of high abiotic stress relative to more benign abiotic conditions (Michalet et al. 2006, Maestre et al. 2009). In the intermediate to severe environments found in the Arctic tundra, both facilitation and competition processes are expected to shape plant distributions and assemblages. However, this expectation remain to be tested with empirical data on species distributions along environmental gradients.

The present study focuses on the tundra of northern Norway, which includes landscapes with varying topography and strong climatic gradients. Within the whole region, the evergreen dwarf shrub *Empetrum nigrum* ssp. *hermaphroditum* (hereafter referred to as *E. hermaphroditum*) is the most abundant plant species (Bråthen et al. 2007a). Its dominance is related to four attributes (Aerts 2010): 1) it is clonal and exhibits a dense pattern of growth and can, thus, pre-empt space locally; 2) it produces phenols that sequester nutrients in the soil and cause low nutrient availability; 3) it produces the allelopathic compound batatasin III, which potentially limits seedling growth and the survival of recruited species in its vicinity; and 4) its leaves are unpalatable to herbivores.

In this study, we investigated the influence of this dominant species on the distribution of subordinate species in plant communities at three spatial scales. Subordinate species may be strongly limited by competitive interactions in addition to climatic constraints and, thus, may exhibit a restricted distribution compared to ranges expected from abiotic factors alone. Conversely, some species may also benefit from the habitat conditions established by the dominant species (Callaway et al. 2002). If this is true, these two types of association (positive and negative) should be apparent in the observed data. We proceeded by comparing the modelled distribution of subordinate species, either including or not including the frequency of the dominant species as a biotic predictor.

Because both biotic and abiotic processes contribute to community assemblages (Adler and HilleRisLambers 2008), variance partitioning approaches can be useful to estimate the relative contributions of biotic interactions and environmental variables (Borcard et al. 1992). We related the proportion of variance explained by the biotic predictor in SDMs to indices used in community ecology: 1) a new index that can be used to compare co-occurrences between species across several scales and 2) values of traits that can describe the functional abilities of plant species during the phase of their life cycle once they are established in the community. To understand how spatial scale affects co-occurrence patterns between the dominant and subordinate species, we applied a study design in which the vegetation was sampled in small plots nested in larger plots (2×2 m, 100×100 m and 750×750 m plots). On this basis, we addressed the following hypotheses:

H_1: when included as an additional predictor into commonly-used topo-climatic SDMs, the frequency of occurrence of *E. hermaphroditum* improves the explained variance and predictive power of topo-climatic SDMs of subordinate tundra species.

H_2: because the response of species to their biotic and abiotic environment depends on their functional attributes, we expect that variation in the direction of the relationship between *E. hermaphroditum* and subordinate species will be related to plant functional traits. The functions associated with traits will provide substantial information on the role of ecological processes associated with the presence *E. hermaphroditum*, such as competition (canopy height, specific leaf area), tolerance to resource depletion (leaf dry matter content, woodiness), or tolerance to grazing (leaf dry matter content).

H_3: the variation in the direction of the statistical relationship between the subordinate tundra species and *E. hermaphroditum* is related to their pattern of diminishing co-occurrence from large to fine scales; that is, when co-occurrence is observed in large plots (750-m plots) but not in the smaller plots (100-m plots or 2-m plots) nested within the former. Because biotic interactions mostly take place at smaller scales, we assume that spatial aggregation or segregation at smaller scales in a homogenous physical environment can be considered as signals of the influence of biotic interactions induced by *E. hermaphroditum*.

Methods

Study area

Finnmark and Troms Counties in Norway form the northern frontier of the European continent, delineated by the Barents Sea towards the north and by birch forests and continuous taiga towards the south. In its western regions, this area is topographically characterised by steep hills running from peaks around 800–1800 m a.s.l., which are often surrounded by glaciers, deep valleys and narrow fjords or the open sea. In the eastern part of Finnmark, the mountain ranges gradually decrease in altitude towards the Barents Sea to plateaus of 300–500 m a.s.l. with abrupt edges interspersed with moderately sloped hills. The relief gradient from west to east has effects on the spatial variability of the local climate. Additionally, in this region there are climatically steep gradients from west to east, as well as from the coast to inland regions caused by the NE Atlantic current, the influence of which gradually declines from west to east. *Empetrum* heaths are widely distributed within this region (Haapasaari 1988), first occurrence dating back 12 500 (cal) yr (Alm 1993).

The current study focuses on the area where the average temperature of the warmest month is $<11^{\circ}$C, corresponding to the beginning of the tundra in northern Norway (Fig. 1). The heaths dominated by *Empetrum nigrum* ssp. *hermaphroditum*, along with *Betula nana* and *Vaccinium* spp., are interspersed with patches of mesic and wet vegetation where herbs (such as *Bistorta vivipara*, *Alchemilla alpina*, *Thalictrum alpinum* and *Viola biflora*) and graminoids (such as *Avenella flexuosa*, *Nardus stricta*, *Carex bigelowii*, *Eriophorum angustifolium*, *Agrostis capillaris* and *Deschampsia cespitosa*) are common species, as is the dwarf shrub *Salix herbacea*.

Environmental predictors

We used monthly average temperature and precipitation grids from the WorldClim climatic maps (Hijmans et al. 2005) at a resolution of 30 arc-seconds (ca 750 m). We combined these grids with our own calculations to develop topo-climatic layers that have recognised strong biological meaning for plants (Zimmermann and Kienast 1999, Guisan and Zimmermann 2000, Zimmermann et al. 2009) and well-established environmental factors that have a clear ecophysiological significance for alpine and tundra flora (Körner 2003): degree-days (DDEG), averaged moisture index of the growth period (June to August, MIND), solar radiation (SRAD), average temperature of the coldest quarter (TCQ), average precipitation of the coldest quarter (PCQ) and intra-annual variation of precipitation (SDPREC).

First, we generated monthly potential global radiation grids (SRAD) consisting of both direct and diffuse components using Kumar's approach (Kumar et al. 1997). Next, we calculated degree-days (DDEG) from monthly average temperatures according to Zimmermann and Kienast (1999). We also calculated potential evapotranspiration grids (Zimmermann and Kienast 1999). The moisture index was calculated as the difference between monthly precipitation and potential evapotranspiration (MIND). Finally, the standard deviation of the monthly precipitation values was calculated (SDPREC). More details on the derivation of topo-climatic predictors can be found in Zimmermann et al. (2007). These environmental predictors were first used to design the environmentally

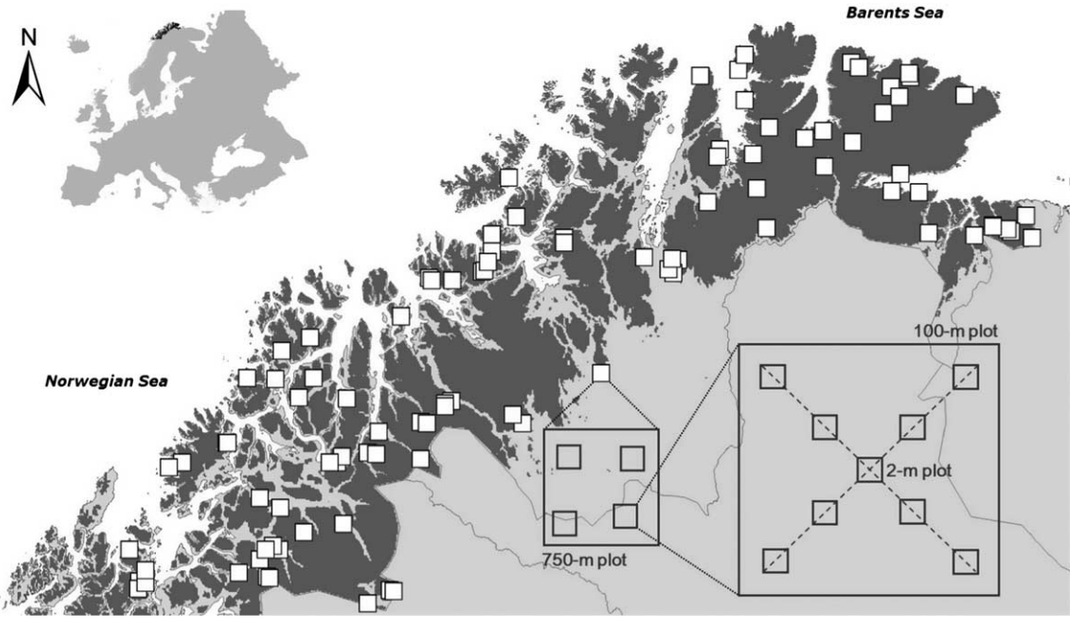


Figure 1. Geographical location of the study area (dark grey) in northern Norway and illustration of the method used to sample vegetation in small plots nested in larger plots (2×2 m, 100×100 m and 750×750 m plots). For better illustration, the plot size on the map has been magnified.

stratified sampling procedures and were then used in the statistical analyses of the SDMs.

Vegetation sampling

Because an inventory of all vascular plant species would have been too time consuming to conduct, we selected 39 plant species representing several growth forms and rarity levels occurring in the low alpine zone of Finnmark. For sampling these species, we stratified our study area using three climatic variables selected for their importance for structuring tundra vegetation: DDEG, MIND and PCQ. All three variables were classified into three equal categories. By combining these three layers, we generated a map representing 25 realised climatic combinations (from the 27 possible) occurring in the alpine zone in the region. We then randomly selected 750×750 m plots (hereafter referred to as 750-m plots) from the map according to an equal-stratified sampling design (Hirzel and Guisan 2002). Each random plot was sub-sampled by randomly selecting four 100×100 m plots (hereafter, 100-m plots), separated from one another by at least 200 m. Within each 100-m plot, we visited 9 plots of 2×2 m (hereafter, 2-m plots) and inventoried the species along a planned route. We set a first 2-m plot at the centre of a 100-m plot (located using a high quality GPS) and recorded all of the target species present. We then moved 25 m to the northeast to sample the second 2-m plot and another 25 m further to the northeast to sample the third 2-m plot before returning to the central point. We performed similar sampling procedures for the three other directions, i.e. northwest, southeast and southwest (Fig. 1). Additionally, we recorded whether the 39 target species were present in the routes between the 2-m plots. During the summer 2008, we visited 98 750-m plots within the inventoried climatic strata. For the following analyses, we considered only those species for which the number of occurrences in the 750-m plots was >20, yielding 34 species plus *E. hermaphroditum* (Supplementary material Table S1).

Testing the importance of a biotic predictor in species distribution models

To address our first hypothesis (H_1), species distribution models were fitted for each species against the set of topo-climatic predictors, with and without the frequency of *E. hermaphroditum* as additional biotic predictor. We calculated the frequency of *E. hermaphroditum* as the number of 2-m plots occupied by the species within the 750-m plots, i.e. a count variable ranging from 0 to 36. This frequency reflects the coverage of *E. hermaphroditum* and, therefore, quantifies the impact this species might have on the presence of other plant species within the 750-m plots. Note that *E. hermaphroditum* occurred within all but two of the 750-m plots, but with varying coverage between plots, which demonstrates its wide distribution in Finnmark.

Modelling methods differ not only based on their performance but also with respect to the biological explanation they can provide. In particular, generalised linear models (GLM, McCullagh and Nelder 1989) provide parameter estimates that are easier to interpret and compare among species than are many other modelling algorithms. We modelled the distribution of the 34 species first with the five topo-climatic predictors alone and then including the frequency of *E. hermaphroditum* as a biotic predictor, using GLMs with a binomial variance function and a logit function as implemented in the R environment (R Development Core Team 2010, Table 1). In addition to linear relationships, we allowed up to second-order polynomials (i.e. quadratic terms) for each predictor. To evaluate the predictive performance of our species distribution models, we used 10-fold cross-validation measured based on the area under the receiver operating characteristic (ROC) curve (AUC, Fielding and Bell 1997). In addition to the AUC, we compared the model fit estimated with the adjusted geometric mean squared improvement R^2 (Cox and Snell 1989, Nagelkerke 1991). This R^2 is rescaled to a maximum of 1 and adjusted for both the number of observations and predictors in the model. We computed the R^2 of the global fit and the AUC both with and without the biotic predictor.

We projected the models over the entire study area. However, because the data on the frequency of *E. hermaphroditum* was available only for the inventoried pixel, we modelled the frequency of *E. hermaphroditum* with a generalised additive model (GAM; Hastie and Tibishirani 1986) with a smoothing function allowing four degrees of freedom and a Poisson distribution using the topo-climatic predictors mentioned earlier. GAMs are sometimes referred to as being data rather than model driven because the data determine the nature of the relationship between the response and the set of explanatory variables, rather than assuming some form of parametric relationship (Guisan et al. 2002). Because our aim was to establish a spatially explicit variable reflecting the spatial frequency of *E. hermaphroditum* as close as possible to the observed frequencies throughout our study area, a GAM was preferred to a GLM for this analysis.

To quantify the respective proportions of the variance explained by the biotic and topo-climatic predictors in the GLMs of each species, we used a variance partitioning approach based on partial regression analyses. This approach allows variance partitioning into four fractions (Borcard et al. 1992; see Randin et al. 2009 for implementation): 1) pure topo-climatic (TC), 2) shared topo-climatic and biotic (TC+BIO), 3) pure biotic (BIO) and 4) unexplained variance. The use of a GLM is required to apply this framework. The adjusted geometric mean squared improvement R^2 was used as an estimator of the explained variance, without adjustment for the number of observations and predictors (see also Zimmermann et al. 2007). Additionally, we extracted from linear GLMs with a binomial variance function and a logit function the sign (positive or negative) of the coefficient of the relationship between the occurrences of each of the 34 species and the biotic predictor alone as well as the associated p-value (from a Wald-z test, Supplementary material Table S2). This allows investigation of the direction of the relationship between the frequency of the dominant species and the occurrence of the subordinate species.

Table 1. Summary of the results regarding the total fit (Tot. R^2), the predictive power (AUC) for the topo-climatic (TC) and biotic (BIO) models and the improvement expressed in percent (Tot. R^2 Imp., AUC Imp.) of the 10 species for which the total fit is the most improved by the frequency of *E. hermaphroditum* (five with a negative relationship in the linear regression with the biotic predictor and five with a positive one). The values of partitioned variance (topo-climatic: R^2 TC, biotic: R^2 BIO, shared: R^2 Sh., and unexplained: R^2 Un.) are also tabulated.

	Sign	R^2 TC	R^2 BIO	R^2 Sh.	R^2 Un.	Tot. R^2 TC	Tot. R^2 BIO	Tot. R^2 Imp.(%)	AUC TC	AUC BIO	AUC Imp.(%)
Veronica alpina	−	0.181	0.165	0.054	0.600	0.228	0.307	34	0.716	0.733	2
Potentilla crantzii	−	0.019	0.131	0.049	0.801	0.116	0.238	105	0.558	0.684	23
Silene acaulis	−	0.208	0.172	0.007	0.614	0.184	0.349	90	0.687	0.732	7
Ranunculus glacialis	−	0.123	0.250	0.273	0.355	0.350	0.574	64	0.775	0.805	4
Arabis alpina	−	0.104	0.382	0.175	0.340	0.287	0.525	83	0.759	0.840	11
Loiseleura procumbens	+	0.194	0.180	0.080	0.546	0.222	0.470	112	0.653	0.738	13
Vaccinium uliginosum	+	0.083	0.161	0.114	0.643	0.363	0.621	71	0.711	0.837	18
Rubus chamaemorus	+	0.044	0.316	0.143	0.497	0.199	0.484	143	0.660	0.815	23
Juncus trifidus	+	0.336	0.186	−0.048	0.526	0.243	0.554	128	0.656	0.840	28
Arctostaphylos alpina	+	0.103	0.216	0.129	0.551	0.257	0.569	122	0.712	0.821	15

Analysing the role of species traits

To address our second hypothesis (H_2), we collected for each of the 34 studied species data on four biological traits from the LEDA database (Knevel et al. 2003) and other sources in the literature (Wijk et al. 2005, Caccianiga et al. 2006, Baptist et al. 2010). We selected leaf dry matter content (LDMC [mg g^{-1}]), woodiness or stem specific density (SSD [g cm^{-3}]), specific leaf area (SLA [mm^2 mg^{-1}]) and canopy height (CH [mm]). These traits can be seen as syndromes of plant species performance related to the phase of the plant life cycle when they are becoming established in the vegetation (Weiher et al. 1999). In particular, under rich nutrient conditions, fast growth is a prerequisite for competitive ability, which is reflected by high SLA and CH values (Weiher et al. 1999, Lavorel et al. 2007). In contrast, under nutrient poor conditions, success can be achieved through resource sequestration ability, which is reflected by a higher LDMC and SSD (Weiher et al. 1999, Lavorel et al. 2007). LDMC is also recognised as being related to environmental stress tolerance, for example, to grazing by large herbivores (Rusch et al. 2009). The analyses were performed on 29 species because trait values were missing for five species. The species were classified into three categories: species that showed a significant positive linear relationship with the biotic predictor (i.e. the frequency of *E. hermaphroditum* see above, +), those that showed a significant negative relationship (−) and finally the species that showed no significant relationship (0). Then we compared the trait values between these three groups using a logistic regression. We also graphically related the variance explained by the biotic predictor to the values of LDMC.

Co-occurrence pattern

To address our third hypothesis (H_3), we calculated the co-occurrence between *E. hermaphroditum* and the 34 subordinate species across our system of nested plots at different scales: 750-m, 100-m and 2-m plots. For each pair of species (S1, S2), we divided the number (N) of plots where both species were present by the number of plots where the rarest of the two species was present. This index ranges from 0 (no co-occurrence) to 1 (always in co-occurrence) as given in eq. 1.

$$\mathrm{Ind}_{\mathrm{coo}} = \frac{N_{(S1 \cap S2)}}{Min(N_{S1}, N_{S2})}, \quad (1)$$

where $N_{(S1 \cap S2)}$ is the number of times species S1 and S2 co-occur, while $Min(N_{S1}, N_{S2})$ is the occurrence frequency of the rarest of the two species.

Because the scale is crucial when analysing co-occurrence patterns, we calculated the reduction in co-occurrences as the spatial scale decreased by taking the difference of the co-occurrence values between the 750-m and 100-m plots and also between the 750-m and 2-m plots. This difference reflects the spatial segregation of the subordinate species and *E. hermaphroditum* within the 750-m plots. If a species frequently co-occurs with the dominant species in the 750-m plot, but not in the nested 100-m or 2-m plot, which is assumed to have similar climatic conditions, we

consider this result to be consistent with a biotic effect (hypothesis 3), although causality can formally only be demonstrated through experiments. If the decrease of co-occurrence is close to average, a biotic effect is less likely. We compared the reduction in co-occurrences between the three groups of species defined above (+, − and 0) using a Kruskal-Wallis test. We also graphically related the variance explained by the biotic predictor to the reduction in co-occurrences.

Results

Testing SDM improvement (H_1)

The SDMs based on topo-climatic variables were improved for most species by including *E. hermaphroditum* as a biotic predictor, but the strength of the improvement was highly variable among the subordinate species. By accounting for the frequency of *E. hermaphroditum*, both the predictive power (on average, +0.04 in the AUC, corresponding to +7%, and up to +0.18, i.e. an increase of 28%, for *Juncus trifidus*, Table 1) and the fit (on average, +0.10 in R^2, corresponding to +50%, and up to +0.31, i.e. an increase of 122% for *Arctostaphylos alpina*, Table 1) were improved. However, for three species, the fit and the predictive power of the model were not improved and were even slightly reduced, and for 9 species, the fit slightly improved, while the predictive power decreased (Supplementary material Table S1). The variance partitioning method indicated that the total explained variance was shared between the topo-climatic and biotic predictors. Twenty-three species had a negative sign, while 11 had a positive sign in the linear relationship between the occurrence of the species and the biotic predictor (Table 1, Supplementary material Table S2 and S3, S4 for details on full models). The GAM based on the frequency of *E. hermaphroditum* explained 39% of the variance. The projections of the models for those species that are highly influenced by the dominant species changed when the frequency of *E. hermaphroditum* was accounted for (Fig. 2).

Testing SDM improvement and subordinate species trait relationships (H_2)

Logistic regression showed that LDMC was the only trait significantly discriminating between the three groups of species, showing either a positive, a negative sign or no significant relationship (+, − and 0) in the linear relationship with the frequency of *E. hermaphroditum* (Table 2). The species with a positive sign had higher LDMCs than species with a negative sign. Also, species for which the variance explained by the biotic predictor was the highest had particularly high or low values of LDMC (Fig. 3). Note that *Rubus chamaemorus* is an outlier of this relationship.

Testing co-occurrence patterns across scales (H_3)

We compared the loss of co-occurrence toward smaller scales between the three groups of species (+, − and 0), but only the reduction from large to small scale showed significant results (Kruskal-Wallis test, 750-m–100-m: $H = 1.91$, $DF = 2$, $p = 0.38$; 750-m–2-m: $H = 19.67$, $DF = 2$, $p < 0.001$; Fig. 3, Supplementary material Fig. S1). The species with a positive sign in the relationship with the biotic predictor co-occurred with *E. hermaphroditum* independent of scale, while the species with a negative sign decreased in co-occurrence from the 750-m to the 2-m plots. Species that either tended to co-occur with *E. hermaphroditum*, or decreased in co-occurrence from the 750-m to the 2-m plots were also those for which the variance explained by the biotic predictor was the highest (Fig. 3).

Discussion

In this study, we used species distribution models (SDM) associated with approaches used in community ecology to improve our understanding of interspecific relationships and their possible inclusion in models. While positive interactions (i.e. facilitation) have been assumed to be more common under harsh conditions (e.g. at high elevations or latitudes) recent experimental studies indicated that co-existing plant species may have both positive and negative effects on each other in cold environments (Choler et al. 2001, Dormann and Brooker 2002). However, drawing similar inferences from single correlative modelling studies is difficult because it is not possible to analyse which of the two interacting species exerts what influence on another (Meier et al. in press), and the potential role of unaccounted environmental factors is difficult to assess. Here, to overcome this problem, we assessed the influence of one largely dominant plant species, *Empetrum hermaphroditum*, on 34 other species in an Arctic tundra ecosystem more specifically. We expected the frequency of *E. hermaphroditum* to improve the explained variance and predictive power of topo-climatic SDMs and to be related to the distribution of most other species (hypothesis 1, H_1), primarily through negative impacts on their distribution. However, we also expected to observe some positive relationships because facilitation is also known to structure plant communities in the Arctic tundra (Dormann and Brooker 2002). Furthermore, the directions of the relationships was expected to be related to functional traits (hypothesis 2, H_2) and co-occurrence patterns (hypothesis 3, H_3) of these species.

In accordance with H_1, the fit and predictive power were improved when *E. hermaphroditum* was added as predictor in the models. Additionally, variation in the direction (i.e. whether negative or positive) of the relationship between *E. hermaphroditum* and subordinate species were linked to one of the functional traits (H_2) and to the co-occurrences patterns (H_3). In the latter case, co-occurrences with *E. hermaphroditum* strongly decreased from the 750-m to the 2-m plots only for subordinate species that were negatively associated with *E. hermaphroditum*. These species typically have LDMC values indicating lower competitive sequestering abilities and higher palatability to herbivores. In contrast, subordinate species for which the co-occurrence was unaffected by the scale were typically positively

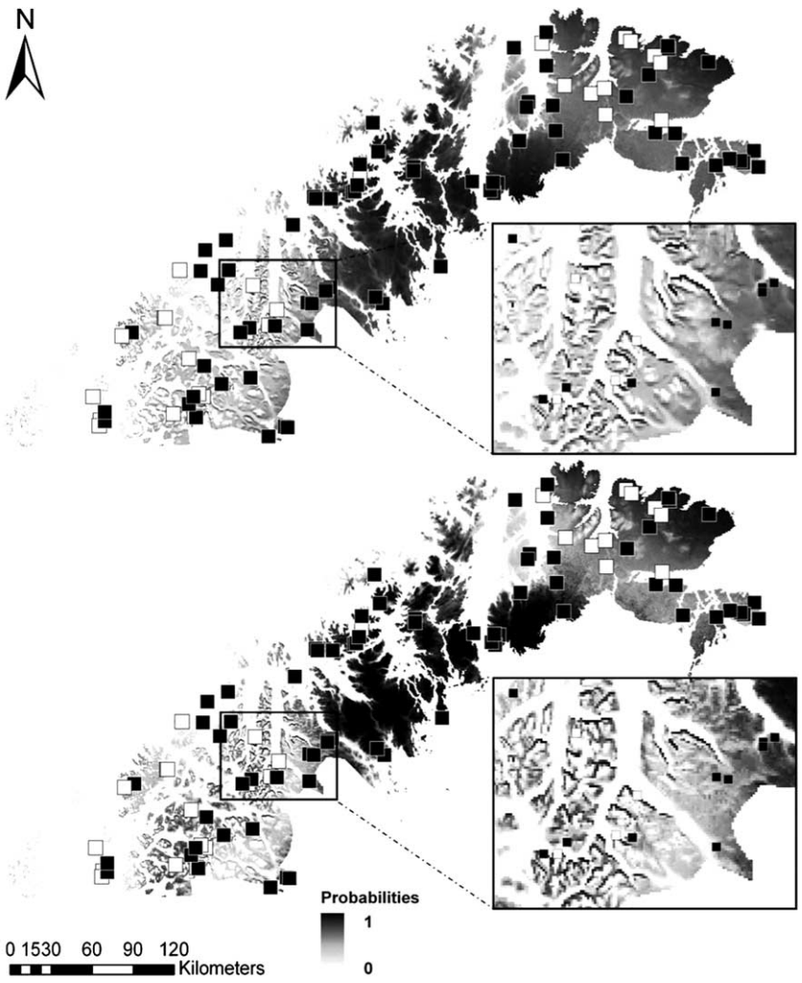


Figure 2. Potential habitat maps of probabilities of occurrence for *Carex bigelowii*, which is positively influenced by *E. hermaphroditum* through a nurse effect. Compared to the projection using only topo-climatic predictors (upper), the fit of the model is improved when the frequency of *E. hermaphroditum* is incorporated (lower). The dark squares indicate observed presences, while the white squares indicate observed absences.

associated with *E. hermaphroditum* and had higher LDMC values. Hence, it appears that *E. hermaphroditum* influences the distributions of other tundra species through both competitive and facilitative effects (as discussed below). This represents a new finding obtained by combining approaches in community ecology with species distribution modelling.

Table 2. Summary of the logistic regression used to discriminate species with a positive relationship with the frequency of *E. hermaphroditum* in the linear regression from the species with a negative relationship, with the p-values (p) highlighted in bold when significant. The leaf dry matter content (LDMC) shows a significant trend, while woodiness (SSD), canopy height (CH) and specific leaf area (SLA) do not.

	Estimate	p
Intercept	−8.266	
SSD	4.120	0.175
CH	5.499	0.632
LDMC	0.015	**0.045**
SLA	0.209	0.166

Species negatively associated with *E. hermaphroditum*

The distributions of most species considered in this study were negatively related to the frequency of *E. hermaphroditum* but were also characterised both by low LDMC and strongly decreased co-occurrences from large to small scales. Therefore, even if these species generally occur under similar climatic conditions (i.e. in the same 750-m plot), they rarely occur in close proximity to *E. hermaphroditum* (i.e. in the same 2-m plot). This result can be interpreted as a competitive effect of *E. hermaphroditum,* either due to exploitation or interference. Additionally, LDMC is a trait that is well recognised as being symptomatic of a plant species' resource use strategy, i.e. the strategy it employs in the fundamental trade-off between rapid assimilation and growth and the efficient conservation of resources within well-protected tissues (Wilson et al. 1999, Garnier et al. 2001, Diaz et al. 2004). By comparing species uptake patterns of soil nitrogen, McKane et al. (2002) showed that the dominance of a species can be related to the species'

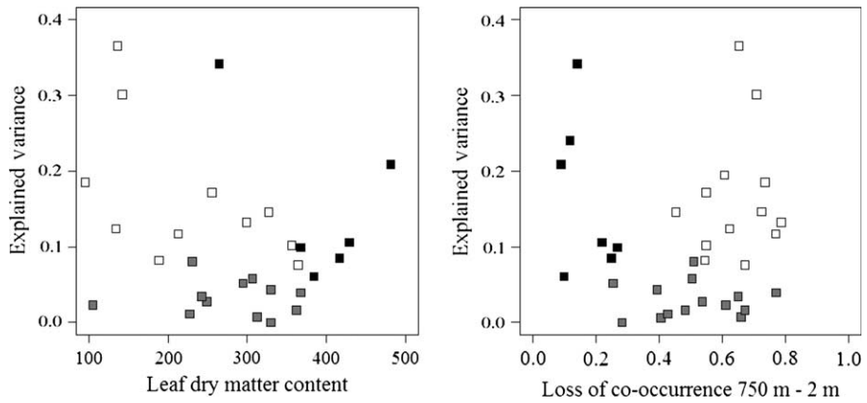


Figure 3. Partial variance of the model explained by the biotic predictor for each of the species plotted against the values of leaf dry matter content (LDMC) (left) and the loss of co-occurrence between the 750 and 2-m plots (right). The small squares represent the three groups of species: species with a significant positive linear relationship with the biotic predictor (black), with a significant negative relationship (white) and finally the species that showed no significant relationship (grey). The black outlier in the upper middle of the left graph is *Rubus chamaemorus*.

ability to exploit the most abundant resource more efficiently. Because nitrogen is a strongly limiting factor in Arctic environments (Parsons et al. 1994) and *E. hermaphroditum* immobilises nutrients (Nilsson 1994), competition for this resource may limit the growth of less competitive species. Species that do not co-occur with *E. hermaphroditum* at the smallest scale investigated are generally palatable grasses and herbaceous dicotyledons. These species have a low LDMC and, thus, do not share with *E. hermaphroditum* the ability to sequester nutrients in their leaves. Rather, they can quickly exploit and release resources, which becomes an advantage in more productive environments. In turn, these species might be considered as weak competitors in nutrient-poor systems. Additionally, *E. hermaphroditum* secretes allelopathic compounds into the soil, which can negatively affect the growth of herbaceous plants (Bråthen et al. 2010). The effects of allelopathic compounds can either be direct or indirect, through a negative influence on soil microbial (Wardle and Nilsson 1997) or mycorrhizal activity (Nilsson 1994), but such allelopathic effects may also lead to humus accumulation and reduced nitrogen availability, in turn negatively impacting plant growth.

We also found that several species with a negative relationship with *E. hermaphroditum* typically grow on rocky outcrops (*Cryptogramma crispa*, *Oxyria digyna*, *Arabis alpina*, *Saxifraga oppositifolia*), which represent a micro-habitat rarely inhabited by *E. hermaphroditum*. These species might be very weak competitors and may, therefore, be outcompeted in areas occupied by *E. hermaphroditum* or other species, thus being restricted to less productive rocky outcrop. Alternatively, the frequency of *E. hermaphroditum* might provide indirect information about the frequency of rocks and boulders, which might be required for the establishment of these species.

Species positively associated with *E. hermaphroditum*

Conversely, woody species and sedges with a high LDMC, which are species that typically adapt to nutrient-poor systems (Stark 2007), had distributions that were positively related to the frequency of *E. hermaphroditum* and co-occurred with *E. hermaphroditum* independent of scale, following our third hypothesis. This limited loss of co-occurrence might be the result of indirect facilitation (Brooker et al. 2008) by *E. hermaphroditum* or of shared environmental filters with *E. hermaphroditum*. That is, the dense cover of *E. hermaphroditum* might influence other species by a nursing effect. Particularly, our results corroborate the facilitative effect observed for *E. hermaphroditum* on *Carex bigelowii* that was shown by Carlsson and Callaghan (1991). However, species co-occurring with *E. hermaphroditum* also have a higher LDMC in general. In this context, it is possible that only species sharing traits reflecting adaptation to nutrient sequestration can tolerate the potential competitive effects of *E. hermaphroditum* through soil resource depletion. Additionally, the vegetative matrix built by *E. hermaphroditum* might also provide other kinds of positive feedback to co-occurring plants, including protection from disturbances, such as desiccation or cold temperatures (Callaway 1995).

Alternatively, the pattern observed for the LDMC values might result from a common environmental filter, such as grazing. Rusch et al. (2009) showed that the most intensely grazed sites on fertile soils possessed plant communities with highly aggregated LDMCs, indicating some environmental filtering through herbivores. The region of Troms and Finnmark is grazed to a considerable extent by reindeer, and species with a high LDMC are less palatable and, thus, will be less affected by grazing (Bråthen et al. 2007). For example, *E. hermaphroditum* is systematically avoided by herbivores (Grime et al. 1988). Other plant characteristics might also be important in this context. For example, even though the LDMC of *Rubus chamaemorus* is not particularly high, the amount of phenolic compounds in its leaves is important (Quested et al. 2005), as this character also limits the effects of grazing (Alonso et al. 2001) and might explain the persistence of this species in the vicinity of *E. hermaphroditum*. Finally, the frequency of *E. hermaphroditum* may affect substrate acidity, so that it facilitates the establishment of some species as for example *Nardus stricta*

or *Juncus trifidus*. A similar pattern was shown for *Vaccinium myrtillus* by Coudun and Gégout (2007). Interestingly, the relationships found in the model were only related to the pattern of a loss of co-occurrence from large to small scales (750 m–2 m), but not from a large to an intermediate scale (750 m–100 m). This finding suggests that the biotic processes leading to facilitation or exclusion occur at a very small scale. Hence, it is only when the coverage of *E. hermaphroditum* extends over a major proportion of the local environment that the species not tolerating *E. hermaphroditum* are excluded (i.e. from the 750-m plot).

Limitations to the approach

One caveat of our approach is that, to limit model complexity and allow proper variance partitioning, we did not account for interactions between variables. On the other hand several studies have documented a change in the direction of interaction from negative to positive between plant species depending on climatic conditions (Brooker and Callaghan 1998, Meier et al. 2010). Hence, even though we found strong relationships in this study, the interactions between plants in a community might, in fact, be more complex. For example, some of the species showing no trend related to the frequency of *E. hermaphroditum* might be positively influenced under certain abiotic conditions, but negatively under others. Furthermore, even though *E. hermaphroditum* as a dominant species is expected to occupy most of the suitable environment, this species appears to also be displaced by other species under some conditions. For example, Nilsson (1994) observed that when the availability of nutrients increases, *E. hermaphroditum* is outcompeted by *Avenella flexuosa*. Other biotic variables, such as the density of reindeer, could also be important in accounting for the frequency of *E. hermaphroditum* through, e.g. grazing (Bråthen et al. 2007) and could improve the fit of the model based on the frequency of *E. hermaphroditum*. Nevertheless, despite the fact that our results are correlative, they suggest new research areas related to the inclusion of biotic interactions in studies of species distributions, which may serve to initiate experimental studies aimed at confirming our results.

Community-based modelling approaches have been proposed to assess interactions between more than one species at a time (Ferrier and Guisan 2006), which could yield further insight about ecological processes underlying SDMs. Comparing individual SDMs incorporating biotic interactions with community-based models accounting for co-occurrence patterns along environmental gradients (Ferrier et al. 2002a, b, Ferrier and Guisan 2006) is one possibility, while making predictions for single species to be assembled later into communities is another (Ferrier and Guisan 2006). However, studies assessing the performance of such community-based models are scarce, and the results are often inconclusive (Baselga and Araújo 2010).

Alternatively, recent studies on functional characteristics of plants have led to the idea that these traits may be linked to the ability of species to face biotic interactions (Gross et al. 2009), especially when distinguishing between species abilities related to rapid acquisition of resources (i.e. exploitative or acquisitive strategy) and long conservation of resources in well protected tissues (i.e. conservative strategy) (Diaz et al. 2004). This study indicates that such traits may help in inferring the role of interactions based on the spatial distribution of species observed in a statistical model. This finding suggests that a functional approach is potentially useful in improving community predictions from single niche-based SDMs, as proposed by the ''predict first, assemble later'' concept of Ferrier and Guisan (2006).

Conclusion

The results of this study are consistent with the existence of both positive and negative interactions taking place between a dominant species and 34 subordinate species. Hence, although competition experiments in the Arctic have generally documented a greater importance of facilitation compared to competition in plant communities, our results suggest that both competition and facilitation may play an important role. Moreover, trait values and co-occurrence patterns observed at small scales suggested processes acting behind the fitted statistical relationships. Hence, by assembling the predictions of subordinate species according to their traits or co-occurrence values, we might improve our predictions of tundra communities compared to a simple stacking of individual species predictions based on environmental predictors alone (Guisan and Rahbek unpubl.). This study highlights the promising results of integrating approaches from community ecology into models of species distribution.

As the Arctic is expected to suffer the greatest degree of warming on the Earth of more than 4°C over the next decades (IPCC 2007), it is essential to develop accurate predictions of the response of species and communities in this region and, therefore, to account for the types of apparent biotic interactions found in this study.

Acknowledgements – We thank all of the people who helped with the field work in Finnmark. This study was supported by the European Commission (ECOCHANGE project, contract nr. FP6 2006 GOCE 036866) and NSF grant Nr. 31003A-125145 (BIOASSEMBLE project).

References

Adler, P. B. and HilleRisLambers, J. 2008. The influence of climate and species composition on the population dynamics of ten prairie forbs. – Ecology 89: 3049–3060.

Aerts, R. 2010. Nitrogen-dependent recovery of subarctic tundra vegetation after simulation of extreme winter warming damage to *Empetrum hermaphroditum*. – Global Change Biol. 16: 1071–1081.

Alm, T. 1993. Øvre Æråsvatn – palynostratigraphy of a 22,000 to 10,000 B.P. lacustrine record on Andøya, northern Norway. – Boreas 22: 171–188.

Alonso, I. et al. 2001. Competition between heather and grasses on Scottish moorlands: interacting effects of nutrient enrichment and grazing regime. – J. Veg. Sci. 12: 249–260.

Anderson, R. P. et al. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. – Oikos 98: 3–16.
Baptist, F. et al. 2010. Direct and indirect control by snow cover over decomposition in alpine tundra along a snowmelt gradient. – Plant Soil. 328: 397–410.
Baselga, A. and Araújo, M. B. 2010. Do community-level models describe community variation effectively? – J. Biogeogr., in press.
Bertness, M. and Callaway, R. M. 1994. Positive interactions in communities. – Trends Ecol. Evol. 9: 191–193.
Borcard, D. et al. 1992. Partialling out the spatial component of ecological variation. – Ecology 73: 1045–1055.
Bråthen, K. A. et al. 2007. Induced shift in ecosystem productivity? Extensive scale effects of abundant large herbivores. – Ecosystems 10: 773–789.
Bråthen, K. A. et al. 2010. Ecosystem disturbance reduces the allelopathic effects of *Empetrum hermaphroditum* humus on tundra plants. – J. Veg. Sci. 21: 786–795.
Brooker, R. W. and Callaghan, T. V. 1998. The balance between positive and negative plant interactions and its relationship to environmental gradients: a model. – Oikos 81: 196–207.
Brooker, R. W. et al. 2008. Facilitation in plant communities: the past, the present, and the future. – J. Ecol. 96: 18–34.
Caccianiga, M. et al. 2006. The functional basis of a primary succession resolved by CSR classification. – Oikos 112: 10–20.
Callaway, R. M. 1995. Positive interactions among plants. – Bot. Rev. 61: 306–349.
Callaway, R. M. et al. 2002. Positive interactions among alpine plants increase with stress. – Nature 417: 844–848.
Carlsson, B. A. and Callaghan, T. V. 1991. Positive plant interactions in tundra vegetation and the importance of shelter. – J. Ecol. 79: 973–983.
Choler, P. et al. 2001. Facilitation and competition on gradients in alpine plant communities. – Ecology 82: 3295–3308.
Condun, C. and Gégout, J. C. 2007. Quantitative prediction of the distribution and abundance of *Vaccinium myrtillus* with climatic and edaphic factors. – J. Veg. Sci. 18: 517–524.
Cox, D. R. and Snell, E. J. 1989. The analysis of binary data. – Chapman and Hall.
Diaz, S. et al. 2004. The plant traits that drive ecosystems: evidence from three continents. – J. Veg. Sci. 15: 295–304.
Dormann, C. F. and Brooker, R. W. 2002. Facilitation and competition in the high Arctic: the importance of the experimental approach. – Acta Oecol. 23: 297–301.
Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. – J. Appl. Ecol. 43: 393–404.
Ferrier, S. et al. 2002a. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. – Biodivers. Conserv. 11: 2275–2307.
Ferrier, S. et al. 2002b. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. – Biodivers. Conserv. 11: 2309–2338.
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: 38–49.
Garnier, E. et al. 2001. Consistency of species ranking based on functional leaf traits. – New Phytol. 152: 69–83.
Gotelli, N. J. and McCabe, D. J. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. – Ecology 83: 2091–2096.
Gotelli, N. J. et al. 2010. Macroecological signals of species interactions in the Danish avifauna. – Proc. Natl Acad. Sci. USA 107: 5030–5035.
Grime, J. P. et al. 1988. Comparative plant ecology: a functional approach to common British species. – Unwin Hyman.
Gross, N. et al. 2009. Linking individual response to biotic interactions with community structure: a trait-based framework. – Funct. Ecol. 23: 1167–1178.
Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Model. 135: 147–186.
Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.
Guisan, A. et al. 2002. Advances in GLM/GAM modeling: from species' distribution to environmental management. – Ecol. Model. 157: 3–4.
Haapasaari, M. 1988. The oligotrophic heath vegetation of northern Fennoscandia and its zonation. – Acta Bot. Fenn. 135: 1–219.
Hastie, T. and Tibishirani, R. 1986. Generalized additive models. – Stat. Sci. 1: 297–318.
Heikkinen, R. K. et al. 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. – Global Ecol. Biogeogr. 16: 754–763.
Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – Int. J. Clim. 25: 1965–1978.
Hirzel, A. and Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. – Ecol. Model. 157: 331–341.
IPCC 2007. Climate change 2007: synthesis report. – Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, IPCC, Geneva.
Keddy, P. et al. 1998. A comparative approach to examine competitive response of 48 wetland plant species. – J. Veg. Sci. 9: 777–786.
Knevel, I. C. et al. 2003. Life-history traits of the northwest European flora: the LEDA database. – J. Veg. Sci. 14: 611–614.
Körner, C. 2003. Alpine plant life, 2nd ed. – Springer.
Kraft, N. J. B. et al. 2008. Functional traits and niche-based tree community assembly in an Amazonian forest. – Science 322: 580–582.
Kumar, L. et al. 1997. Modelling topographic variation in solar radiation in a GIS environment. – Int. J. Geogr. Inf. Sci. 11: 475–497.
Lavorel, S. et al. 1997. Plant functional classifications: from general groups to specific groups based on response to disturbance. – Trends Ecol. Evol. 12: 474–478.
Lavorel, S. et al. 2007. Plant functional types: are we getting any closer to the Holy Grail? – In: Canadell, J. G. et al. (eds), Terrestrial ecosystems in a changing world. Springer, pp. 149–164.
Leathwick, J. R. and Austin, M. P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. – Ecology 82: 2560–2573.
Lortie, C. J. et al. 2004. Rethinking plant community theory. – Oikos 107: 433–438.
Maestre, F. T. et al. 2009. Refining the stress-gradient hypothesis for competition and facilitation in plant communities. – J. Ecol. 97: 199–205.
McCullagh, P. and Nelder, J. A. 1989. Generalized linear models, 2nd ed. – Chapman and Hall.
McKane, R. B. et al. 2002. Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra. – Nature 415: 68–71.
Meier, E. S. et al. 2010. Biotic and abiotic variables show little redundancy in explaining tree species distribution. – Ecography 33: 1038–1048.

Meier, E. S. et al. in press. Competition patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of *Fagus sylvatica*. – J. Biogeogr.
Michalet, R. et al. 2006. Do biotic interactions shape both sides of the humped-back model of species richness in plant communities? – Ecol. Lett. 9: 767–773.
Nagelkerke, N. J. D. 1991. A note on a general definition of the coefficient of determination. – Biometrika 78: 691.
Nilsson, M. C. 1994. Separation of allelopathy and resource competition by the boreal dwarf shrub *Empetrum hermaphroditum* Hagerup. – Oecologia 98: 1–7.
Parsons, A. N. 1994. Growth responses of four sub-Arctic dwarf shrubs to simulated environmental change. – J. Ecol. 82: 307–318.
Pulliam, H. R. 2000. On the relationship between niche and distribution. – Ecol. Lett. 3: 349–361.
Quested, H. M. et al. 2005. The impact of hemiparasitic plant litter on decomposition: direct, seasonal and litter mixing effects. – J. Ecol. 93: 87–98.
R Development Core Team 2010. R: a language and environment for statistical computing. – R Foundation for Statistical Computing.
Randin, C. F. et al. 2009. Land use improves spatial predictions of mountain plant abundance but not presence–absence. – J. Veg. Sci. 20: 996–1008.
Rusch, G. M. et al. 2009. Plant traits link hypothesis about resource-use and response to herbivory. – Basic Appl. Ecol. 10: 466–474.
Stark, S. 2007. Nutrient cycling in the tundra. – In: Marschner, P. and Rengel, Z. (eds), Nutrient cycling in terrestrial ecosystems. Soil Biol. 10: 309–331.

Download the Supplementary material as file E6386 from <www.oikos.ekol.lu.se/appendix>.

Tilman, D. 1994. Competition and biodiversity in spatially structured habitats. – Ecology 75: 2–16.
Wardle, D. A. and Nilsson, M. C. 1997. Microbe-plant competition, allelopathy and arctic plants. – Oecologia 109: 291–293.
Wardle, D. A. et al. 1998. Can comparative approaches based on plant ecophysiological traits predict the nature of biotic interactions and individual plant species effects in ecosystems? – J. Ecol. 86: 405–420.
Weiher, E. and Keddy, P. 1999. The scope and goals of research on assembly rules. – In: Weiher, E. and Keddy, P. (eds), Ecological assembly rules: perspectives, advances, retreats. Cambridge Univ. Press, pp. 1–20.
Weiher, E. et al. 1999. Challenging Theophrastus: a common core list of plant traits for functional ecology. – J. Veg. Sci. 10: 609–620.
Wijk, M. T. V. et al. 2005. Tight coupling between leaf area index and foliage N content in Arctic plant communities. – Oecologia 142: 421–427.
Wilson, P. J. et al. 1999. Specific leaf area and leaf dry matter content as alternative predictors of plant strategies. – New Phytol. 143: 155–162.
Zimmermann, N. E. and Kienast, F. 1999. Predictive mapping of alpine grasslands in Switzerland: species versus community approach. – J. Veg. Sci. 10: 469–482.
Zimmermann, N. E. et al. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. – J. Appl. Ecol. 44: 1057–1067.
Zimmermann, N. E. et al. 2009. Climatic extremes improve predictions of spatial patterns of tree species. – Proc. Natl Acad. Sci. USA 106: 19723–19728.

ANNEX 2

The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients

Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2013) **22**, 52–63

A Journal of Macroecology

Global Ecology and Biogeography

RESEARCH PAPER

The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients

Julien Pottier[1]*, Anne Dubuis[1], Loïc Pellissier[1], Luigi Maiorano[1], Leila Rossier[1], Christophe F. Randin[2], Pascal Vittoz[1] and Antoine Guisan[1]

[1]*Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland,* [2]*Botanisches Institut der Universität Basel, Basel, Switzerland*

ABSTRACT

Aim Climatic niche modelling of species and community distributions implicitly assumes strong and constant climatic determinism across geographical space. We tested this assumption by assessing how stacked-species distribution models (S-SDMs) perform for predicting plant species assemblages along elevation gradients.

Location The western Swiss Alps.

Methods Using robust presence–absence data, we first assessed the ability of topo-climatic S-SDMs to predict plant assemblages in a study area encompassing a 2800-m wide elevation gradient. We then assessed the relationships among several evaluation metrics and trait-based tests of community assembly rules.

Results The standard errors of individual SDMs decreased significantly towards higher elevations. Overall, the S-SDM overpredicted far more than they underpredicted richness and could not reproduce the humpback curve along elevation. Overprediction was greater at low and mid-range elevations in absolute values but greater at high elevations when standardized by the actual richness. Looking at species composition, overall prediction success, kappa and specificity increased with increasing elevation, while the Jaccard index and sensitivity decreased. The best overall evaluation – as driven by specificity – occurred at high elevation where species assemblages were shown to be subject to significant environmental filtering of small plants. In contrast, the decreased overall accuracy in the lowlands was associated with functional patterns representing any type of assembly rule (environmental filtering, limiting similarity or null assembly).

Main conclusions We provide a thorough evaluation of S-SDM emphasizing the need to carefully interpret standard evaluation metrics, which reflect different aspects of assemblage predictions. We further reported interesting patterns of change in S-SDM errors with changes in assembly rules along elevation. Yet, significant levels of assemblage prediction errors occurred throughout the gradient, calling for further improvement of SDMs, e.g. by adding key environmental filters that act at fine scales and developing approaches to account for variations in the influence of predictors along environmental gradients.

Keywords

Assembly rules, community assembly, elevation gradient, plant traits, species distribution models (SDMs), species diversity.

*Correspondence: Julien Pottier, INRA, Grassland Ecosystem Research (UR874), 5 Chemin de Beaulieu F-63069, Clermont Ferrand Cedex 2, France.
E-mail: julien.pottier@clermont.inra.fr

DOI: 10.1111/j.1466-8238.2012.00790.x

INTRODUCTION

Understanding the role of the environment in shaping species diversity patterns has long motivated ecological and biogeographical research on local to global scales. In recent years, this research has greatly benefited from the development of large species occurrence databases and from conceptual and technical advances in niche based species distribution models (SDMs; Guisan & Thuiller, 2005; Elith & Leathwick, 2009). SDMs are used to predict the potential distribution of species in space and time by relating observed occurrence or abundance patterns to a set of environmental variables (Guisan & Thuiller, 2005). When most species pertaining to a given geographical species pool are considered, such as the whole flora of a given area, different structural or compositional aspects of species assemblages can be predicted. For example, several studies have used stacked-SDMs (S-SDMs) to predict current and future distributions of species richness (e.g. Guisan & Theurillat, 2000; Feria & Peterson, 2002; Algar *et al.*, 2009; Newbold *et al.*, 2009; Pineda & Lobo, 2009) or turnover (Thuiller *et al.*, 2005; Maiorano *et al.*, 2011) under various scenarios of climate change.

Geographical distributions of species are constrained by strict eco-physiological requirements and various other factors, including dispersal processes, positive and negative biotic interactions as well as anthropogenic and geomorphic perturbations (Soberón, 2007). Thus, SDMs are often assumed to fit spatial realizations of the environmental niche of the studied species (Araújo & Guisan, 2006). However, there are still ambiguities over the components of the realized niches that are estimated in climate-based SDMs (Elith & Leathwick, 2009). In addition, there is a large amount of evidence indicating that the relative importance of species distribution/assembly drivers are not constant over space and time or along productivity gradients (Michalet *et al.*, 2006) and that these factors can mutually influence each other along these trajectories (Agrawal *et al.*, 2007). Finally, little is known about how SDMs are affected by such variations of species distribution/assembly drivers.

Despite an increasing use of climate-based S-SDMs and promising associated perspectives, few studies have evaluated the models' predictive accuracy (Feria & Peterson, 2002; Algar *et al.*, 2009; Newbold *et al.*, 2009; Pineda & Lobo, 2009; Trotta-Moreu & Lobo, 2010; Dubuis *et al.*, 2011). To our knowledge, no study has assessed whether the performance of S-SDMs for predicting communities is constant throughout space or along environmental gradients. Such an evaluation is important because S-SDMs represent crucial tools for assessing how future species assemblages will look under future climatic conditions (Ferrier & Guisan, 2006; Guisan & Rahbek, 2011).

In harsh or stressful climatic conditions, community assembly is commonly driven by environmental filtering, which permits only those species sharing the appropriate physiological, behavioural and/or ecological attributes required to survive in the local climate to coexist (Weiher *et al.*, 1999). In this case, climate directly affects the physiology of the species and represents an important filter of community assembly. In addition, climate influences the nature and strength of community- and ecosystem-level processes, which also determine the assembly and distribution of species. For example, it has been reported that low summer temperatures are associated with an increase in the occurrence of facilitative effects among plants (Callaway *et al.*, 2002). Alternatively, the importance of species interactions has been shown to be reduced in alpine ecosystems (Mitchell *et al.*, 2009). Because precipitation and frost partially control geomorphological processes (including erosion), these factors are also responsible for increased disturbance regimes. These examples emphasize climate as a strong direct or indirect determinant of species distribution and assembly, especially in harsh conditions, where more accurate SDM and S-SDM predictions should thus be expected.

In milder climates, ecosystems are more productive, and biotic interactions, such as competition, may be more important than purely abiotic effects for shaping both species distributions and communities (Grime, 2001). In the context of the realized niche, a typical expected consequence of competitive exclusion is a restriction in the occupation of the fundamental niche space. Therefore, predictions based on species responses modelled from distribution data should also account for the role of biotic interactions. In addition, climatically mild habitats are also noticeably affected by intense and diverse human activities, which locally modify abiotic and biotic factors and/or enhance the stochasticity driven by the interplay between disturbance regimes and dispersal events. This impact leads to situations in which the species are likely not at equilibrium with the climate (Araújo & Pearson, 2005). Among locations sharing similar climatic constraints, the same species may encounter very different biotic and abiotic constraints, which may cause the species to exhibit different performances in terms of establishment, growth and fitness. In such conditions, climatic factors may make less of a contribution to the distribution and assembly of the species, and climatic niche models would accordingly fit weaker species responses and produce predictions that are less accurate. As a result, the S-SDM predictions of such communities are likely to be less accurate.

In summary, when stacking climate-based SDMs to predict species assemblages, and provided that the most relevant environmental predictors are available, one could expect the following (Fig. 1): (1) H1 – the accuracy of assemblage predictions increases when moving from productive and mild conditions towards climatically stressful habitats; (2) H2A – the best S-SDM performance in the harshest habitats is primarily associated with strong environmental filtering; and (3), as a corollary, H2B – the inaccurate assemblage predictions in milder habitats are associated with a larger spectrum of constraints driving the local assembly of species.

The objective of this study was to conduct a thorough evaluation of S-SDMs. We tested hypotheses H1 and H2 using elevation as a gradient of climate stress and habitat productivity. We used a large vegetation dataset derived from a robust sampling at a fine resolution, covering the full elevation range of the western Swiss Alps. We reconstructed plant communities by stacking predictions from individual SDMs based on high-resolution topo-climatic predictors and evaluated the deviation between

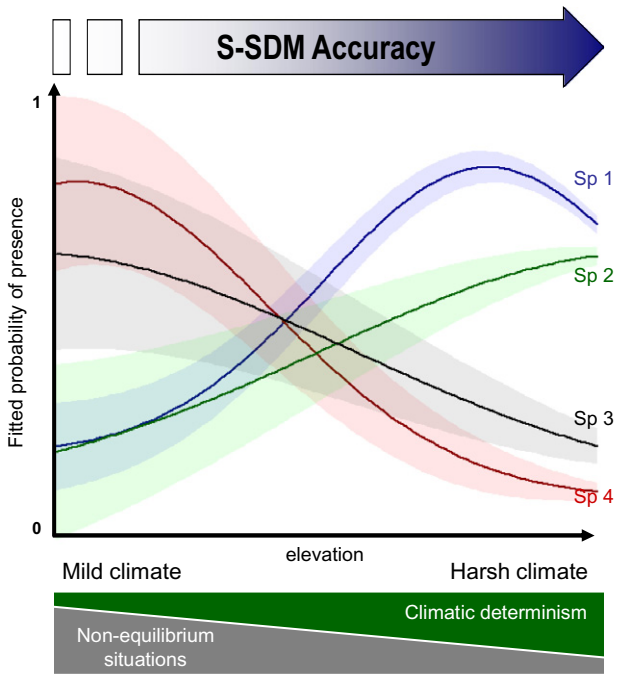


Figure 1 Hypothetical explanation and expected variation in the performance of stacked-species distribution models (S-SDMs) along elevation. At high elevations, the climate is the main determinant of species assembly and species responses to topo-climatic environment are precisely fitted (i.e. tight confidence intervals). Stacked-species distributions are thus expected to provide accurate predictions of species assemblages. In contrast, at low elevations, species distribution is not always at equilibrium with the climate and is rather caused by a variety of assembly rules (possibly mediated by human activities) while climatic conditions remain the same. Fitted species responses to the climate are weak (i.e. large confidence intervals) and the accuracy of the S-SDMs is reduced.

the predicted and observed species assemblages using an independent dataset. Finally, we assessed whether the performance of S-SDMs changed with the spatial variations in the mean and dispersion patterns of the plant functional traits, which are used to infer the constraints that drive the assembly of communities (assembly rules in their broader definition, following Keddy, 1992). Trait convergence was interpreted as a signature of environmental filtering, implying that only species that share common ecological abilities to face local environmental conditions can coexist. Conversely, trait divergence was interpreted as a signature that two competing species cannot coexist unless they exhibit limited similarity in their ecological requirements.

MATERIALS AND METHODS

Study area, species and trait data

The study area covered approximately 700 km^2 of a mountainous region in the western Swiss Alps (Fig. 2a). This region is characterized by a large elevation gradient, ranging from 375 to 3210 m. Human activities are more extensive at the lower elevations (Fig. 2b).

A total of 613 2 m × 2 m vegetation plots were exhaustively surveyed, and the presence–absence of each species was registered. Each plot was selected following a random-stratified sampling design (Hirzel & Guisan, 2002) limited to open and non-woody vegetation units; the stratification was based on elevation, slope and aspect. This presence–absence dataset was used for the SDM calibration. An additional set of 298 plots was identically surveyed to evaluate S-SDMs and to test the relationship between the functional trait patterns of communities and S-SDM errors. The abundance of each species was additionally recorded in all of the plots following an ordinal scale (0, absent; 1, ≤ 1%; 2, 1–5%; 3, 5–12.5%; 4, 12.5–25%; 5, 25–50%; 6, 50–75%; 7, > 75%).

Three plant traits related to the plant persistence phase (Westoby *et al.*, 2002) were investigated: canopy height (CH), specific leaf area (SLA) and leaf dry matter content (LDMC). CH is associated with the plant's ability to compete for light, while SLA and LDMC are related to the plant's ability to capture, use and release resources from/to the plant's environment. These traits were measured within the same study area following Cornelissen *et al.* (2003) over a minimum of 10 individuals per species picked from the available 911 plots (613 calibration plots plus 298 evaluation plots).

Species distribution modelling

We modelled the spatial distribution of 211 species (the species with more than 20 occurrences across the 613 plots). This set of species was considered as the 'landscape species pool'. We used five topo-climatic explanatory variables to calibrate the SDMs: growing degree days (with a 0°C threshold); moisture index over the growing season (average values for June to August in mm day^{-1}); slope (in degrees); topographic position (an integrated and unitless measure of topographic exposure; Zimmermann *et al.*, 2007); and the annual sum of radiation (in kJ m^{-2} year^{-1}). The spatial resolution of the predictor maps was 25 m × 25 m so that the models could capture most of the small-scale variations of the climatic factors in the mountainous areas. These variables are well recognized as being of great ecophysiological importance for plant species in mountain regions and have already been successfully used (Randin *et al.*, 2010). Single-species models were performed following an ensemble forecasting approach based on the weighted average of multimodel predictions weighted by their respective area under the curve (AUC) of a receiver operating characteristic plot (ROC), as proposed by Araújo & New (2007). We considered the following eight modelling techniques: generalized linear model (GLM); generalized additive model (GAM); generalized boosted regression models (BRT); multivariate adaptive regression splines (MARS); random forest (RF); classification tree analysis (CTA); and surface range envelope (SRE). Species distribution modelling was performed using the BIOMOD platform (Thuiller *et al.*, 2009) for R (R Development Core Team, 2011). The spatial and temporal autocorrelation of the residuals of the SDMs was assessed between the training and the evaluation

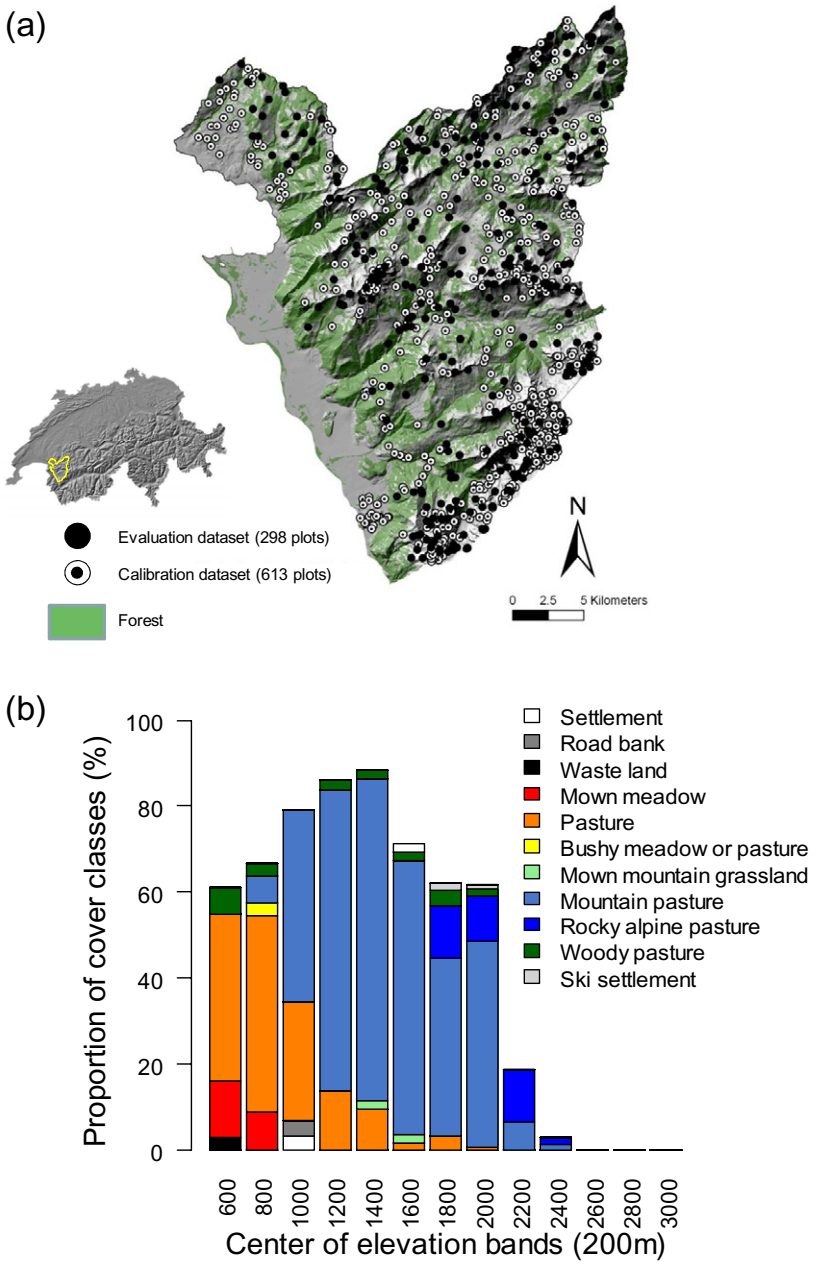


Figure 2 (a) Maps of plot locations, for the calibration and evaluation datasets, within the study area located in the western Swiss Alps. (b) Human land-use distribution in elevation bands for the same study area. Land-cover classes not related to human activities were removed. As observed at a more global scale (Nogués-Bravo *et al.*, 2008), land-use is more intense at the lowest elevations and more diverse.

datasets based on neighbourhood plots and Moran's *I* coefficient (see Appendix S1 in Supporting Information).

We further estimated the confidence intervals of the SDM predictions at each elevation band. To do so, we fitted GAMs for each species and predicted the probabilities of occurrence for each calibration plot as well as their associated confidence interval (following the Bayesian confidence interval computed with the 'mgcv' package for R, according to Wood, 2006). Finally, we simply tracked their variation along the elevation gradient. Thus, the GAMs were fitted with the same set of topo-climatic variables as the ensemble models; however, the GAMs used a randomly selected subset of plots to obtain a homogeneous distribution of plots along the elevation gradient. This subsampling strategy was designed to avoid any potential bias in the confidence interval estimates caused by uneven numbers of observations among the elevation classes. Because there is a random component in this subsampling procedure, the sampling and model fitting were repeated 9999 times. The mean confidence intervals (over sampling repetitions) of the predicted probabilities for each species were retrieved for each plot, and the average for each elevation band was calculated. For each species, the results included only the elevation bands where the species was observed at least once in the calibration dataset.

Evaluating stacked-SDMs

Using maps of the topo-climatic variables that were considered, the SDMs were projected over the geographical space of the

Table 1 Metrics used to evaluate the prediction accuracy of stacked species distribution models (S-SDMs) along elevation in the western Swiss Alps.

Evaluation metric of species assemblage predictions	Formula
Species richness errors	Predicted – observed species richness
Assemblage prediction success	$\frac{a+d}{N}$
Assemblage kappa	$\frac{\left(\frac{a+d}{N}\right)-\frac{(a+b)(a+c)+(c+d)(d+b)}{N^2}}{1-\frac{(a+b)(a+c)+(c+d)(d+b)}{N^2}}$
Assemblage specificity	$\frac{d}{b+d}$
Assemblage sensitivity	$\frac{a}{a+c}$
Jaccard	$\frac{a}{a+b+c}$

Most of these metrics rely on confusion matrices in which species (N = 211) from the landscape species pool are classified into: *a*, species that are both observed and predicted as being present; *b*, species that are observed as absent but predicted as present; *c*, species that are observed as present but predicted as absent; and *d*, species that are both observed and predicted as absent. Note that such confusion matrices and several of the associated metrics are the same as those used to evaluate predictions from single SDMs but applied to species assemblages (counting species instead of occurrences). To avoid confusion, we added the term 'assemblage' prior to each metric name.

study area. Using these maps, we obtained the predicted probabilities of presence for each species and each pixel over the entire study area, including the 298 plots of our evaluation dataset.

Evaluating S-SDM performance for predicting assemblage composition and species richness requires species presence–absence predictions, while SDMs provide probabilities. Because the choice of a threshold for classifying the predicted probabilities has been shown to have an impact on the accuracy of S-SDM predictions (Pineda & Lobo, 2009), we used a novel approach that randomly generated binary predictions from a binomial distribution with the predicted probability of the species (Dubuis *et al.*, 2011). We performed this procedure 9999 times and, at each trial, calculated six evaluation metrics (Table 1) for each plot: (1) species richness errors, i.e. the difference between the observed and predicted species richness; (2) assemblage prediction success, i.e. the proportion of correct predictions; (3) assemblage kappa, i.e. the proportion of specific agreement; (4) assemblage specificity, i.e. the proportion of absences that were correctly predicted; (5) assemblage sensitivity, i.e. the proportion of presences that were correctly predicted; and (6) the Jaccard index, a widely used metric of community similarity.

To test our first hypothesis (H1: S-SDM performance increases with increasing elevation as a result of tighter confidence intervals of the modelled species' responses to the climate), we used nonparametric generalized additive modelling to fit the relationship between these six metrics and elevation, also accounting for nonlinear trends.

Understanding prediction errors

We tested our second general hypothesis (H2: the variation in S-SDM performance is related to changes in the assembly rules constraining the communities) by testing the relationship between the S-SDM performance and functional patterns of plant communities towards high (H2A) and low (H2B) elevations.

Functional patterns were assessed in the 298 evaluation plots using the three traits measured at the species level (CH, SLA and LDMC). The traits could only be measured in the field for the 254 most frequent species (out of the 771 identified). Therefore, we computed community aggregated trait values and functional α-diversity for the plots where trait-assigned species accounted for more than 80% of the vegetation cover (Pakeman & Quested, 2007). This procedure restricted the number of plots that could be used for functional pattern analyses, although the procedure conserved 85% of the plots in the original evaluation dataset (256 plots). The community aggregated trait values ($Trait_{CA}$) consisted of trait averaging of the species found in a given plot weighted by their estimated cover. This measure has been shown to provide useful insight to community dynamics and ecosystem functioning (Garnier *et al.*, 2004). We calculated the α-Rao index following de Bello *et al.* (2010). Doing so, we calculated the sum of functional dissimilarities between all of the possible pairs of species, which were weighted by the product of their estimated cover. Therefore α-Rao is strictly an index of functional evenness. We first calculated the modified α-Rao index for a three-dimensional functional space, considering all of the measured traits together, and then calculated the same index for three one-dimensional functional spaces, considering each trait separately.

To infer assembly rules and discriminate between limiting similarity and environmental filtering, we tested for any deviation of the observed Rao index in the evaluation plots from a null distribution of trait values (divergence or convergence). For a given plot, the null distribution was generated by randomly re-assigning (9999 times) the trait values of the co-occurring species from the trait matrix composed of the entire pool of species (211 species). The actual values of each species for the three traits were kept as a set to conserve the fundamental trade-offs among traits. This simple randomization procedure allowed for testing the null hypothesis that trait values are randomly assembled, while keeping species richness patterns, species frequency patterns and abundance patterns within the plots constant. Then, we computed a standardized effect size (SES) for the Rao values, as proposed by Gotelli & McCabe (2002), as follows:

$$\text{SES-Rao} = \frac{\text{Rao}_{\text{obs}} - \overline{\text{Rao}}_{\text{sim}}}{\sigma\text{Rao}_{\text{sim}}}$$

where Rao_{obs} represents the observed Rao value, Rao_{sim} is the average of the Rao values simulated using the null model, and

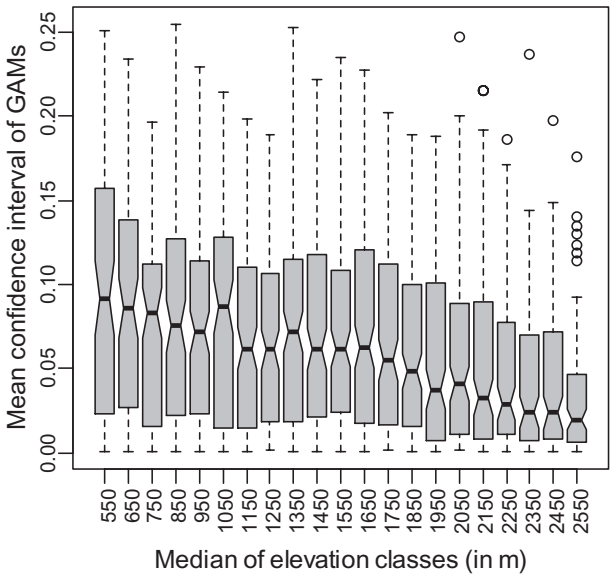


Figure 3 Change in the accuracy of fitted species response to the topo-climatic factors along elevation. This was assessed using generalised additive models (GAMs) and the confidence interval of predicted probabilities for the calibration plots. Each sample represents the mean confidence interval of a given species for a given elevation band.

σRao_{sim} is the variance. Positive values indicate trait divergence, and negative values indicate trait convergence. This standardization of the Rao index allows for a better comparison of the functional evenness across different plant communities (Gotelli & McCabe, 2002).

Finally, we tested the relationship between $Trait_{CA}$, $Trait_{SES\text{-}Rao}$ and metrics of S-SDM performance in the plots using GAMs to facilitate the detection of nonlinear trends.

RESULTS

SDMs were successfully calibrated for the 211 dominant species in the study area and showed fair to good prediction accuracy (all of the AUCs were greater than 0.7; Araújo *et al.*, 2005). For the large majority of the species, no spatial autocorrelation was observed in the residuals between the calibration and evaluation datasets, and the correlations remained very low even when significant (average Moran's $I = 0.143$ for the shortest distances separating the calibration and evaluation plots; see Appendix S1). We therefore considered the second dataset of 298 plots as valid for the S-SDM evaluation.

The modelled species responses using GAMs were not consistently well adjusted along the elevation gradient. In particular, smaller standard prediction errors were observed at high elevations than at middle or low elevations (Fig. 3).

Using S-SDMs to predict species assemblages in our evaluation dataset resulted in a significant relationship between the predicted and observed species richness ($r = 0.45$, $P < 0.001$; Fig. 4c). The slope estimate of this relationship was 0.23, the standard error was 0.03 and the intercept estimate was 26.52 with a standard error of 0.73. The mean species richness error in the overall dataset was 7.30 (SE = 0.58), the mean assemblage prediction success was 0.78 (SE = 0.003), the mean assemblage kappa was 0.72 (SE = 0.004), the mean assemblage specificity was 0.85 (SE = 0.001), the mean assemblage sensitivity was 0.23 (SE = 0.004) and the mean Jaccard index was 0.11 (SE = 0.003).

The predicted species richness pattern in the evaluation dataset (Fig. 4b) did not reproduce the observed hump-shaped curve with a peak of diversity at 1500–1700 m (Fig. 4a). Instead, the S-SDM predicted a progressive and slow (compared with the observed) decrease of species richness as the elevation of the plots increased (Fig. 4b). In addition, the different evaluation metrics reported here showed significant variations along the elevation gradient (Fig. 4d–i). More specifically, we observed a larger overprediction of species richness at low elevations than at mid-range and high elevations, and a larger overprediction was observed at high elevations than at mid-range elevations (Fig. 4d, Appendix S2). Next, we observed very similar nonlinear trends of assemblage prediction success (Fig. 4e) and assemblage kappa (Fig. 4f), with no significant variation from low to mid-range elevations and a strong increase from mid-range to high elevations (Appendix S2). The specificity and sensitivity showed significant linear variations along elevation, positive for the specificity (linear regression $R^2 = 0.52$, $P < 0.001$; Fig. 4g) and negative for the sensitivity (linear regression $R^2 = 0.41$, $P < 0.001$; Fig. 4h). Conversely, the Jaccard index showed a decrease towards higher elevations (Fig. 4i, Appendix S2). The variation of most of the evaluation metrics was strongly correlated with the observed plant species richness, except the assemblage specificity and sensitivity (Fig. S3.1 in Appendix S3). Species richness errors standardized by the observed species richness were higher by far at high elevations compared with mid-range or low elevations. The assemblage prediction success, kappa and specificity, standardized by the observed species richness, confirmed the increase of S-SDM performance with elevation. The relative sensitivity also increased with elevation, whereas the Jaccard index showed a U-curve-like trend (Fig. S3.2 of Appendix S3). The absolute sensitivity (and to some extent the Jaccard index) and specificity were strongly and positively correlated with the predicted species richness (Fig. S3.1 of Appendix S3): the more species the S-SDM predicts, the better it predicts true presences, and conversely the fewer species it predicts, the poorer it predicts true absences.

We only report here on the assemblage prediction success, assemblage specificity and assemblage sensitivity for the analyses that tracked the association between the assembly rules and variation in the S-SDM accuracy. The results with the other metrics provided complementary support to our conclusion or non-significant trends (Appendix S4). The assemblage prediction success was only significantly related to the community aggregated canopy height (CH_{CA}; $R^2 = 0.49$; Fig. 5b). We observed a significant decrease ($R^2 = 0.21$; Fig. 5c) of assemblage specificity with the deviation in the Rao index of canopy height values compared with the null expectation (measured as the SES of Rao: $CH_{SES\text{-}Rao}$), and the specificity decreased with the community aggregated canopy height (CH_{CA}; $R^2 = 0.49$; Fig. 5d). The

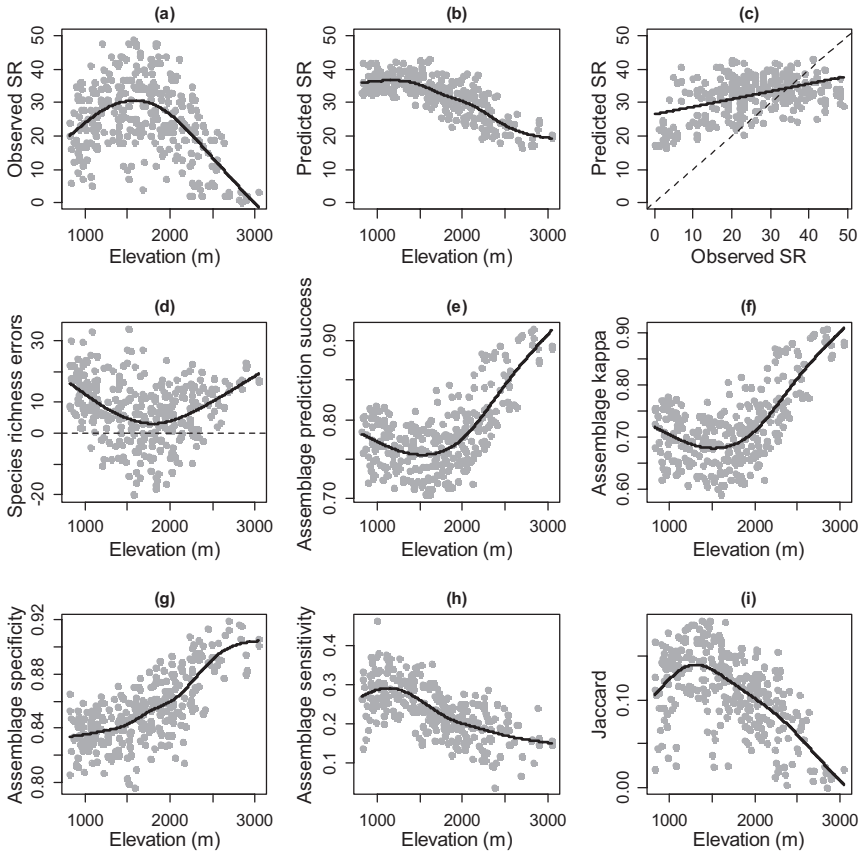


Figure 4 The evaluation of stacked-species distribution models (S-SDMs) based on topo-climatic predictors. Calibration of topo-climatic niche models of the species was based on 613 plots, while evaluation of S-SDMs was based on 298 different evaluation plots (see Supporting Information for a test of independence between the calibration and evaluation datasets). Solid lines represent significant trends fitted with generalized additive models.

assemblage sensitivity increased with $CH_{SES\text{-}Rao}$ (R^2 = 0.13; Fig. 5e) and CH_{CA} (R^2 = 0.21; Fig. 5g). The observed patterns revealed that the best prediction success and specificity, the worst sensitivity and the highest elevations were almost exclusively associated with the significant convergence of CH. On the contrary, the worst prediction success and specificity, the best sensitivity and the lowest elevations were associated equally with the convergence, divergence and null distribution of CH. Regarding SLA and LDMC, we observed significant, although weak, positive relationships between $SLA_{SES\text{-}Rao}$ (SLA for specific leaf area) and $LDMC_{SES\text{-}Rao}$ (LDMC for leaf dry matter content) with assemblage specificity and negative relationships with assemblage specificity (Appendix S4). $LDMC_{CA}$ showed a significant but hardly interpretable inverse-parabolic (decreasing then increasing) trend with the assemblage specificity and a parabolic (increasing then decreasing) trend with the assemblage sensitivity. SLA_{CA} showed a significantly negative relationship with the assemblage specificity and a positive relationship with the sensitivity (Appendix S4).

DISCUSSION

Using S-SDMs, we documented detailed variations in the accuracy of assemblage predictions along a 2700-m wide gradient, which is considered partially as a surrogate of a stress–productivity gradient. Such an evaluation had not previously been performed, despite its crucial importance for assessing the reliability of S-SDMs as a tool to derive scenarios of biodiversity responses to global change along large environmental gradients (Thuiller *et al.*, 2005; Engler *et al.*, 2011). We further showed that the S-SDM accuracy changed with community assembly rules, using analyses of functional trait patterns at the community level. The exact causes behind these observations require careful interpretation and are discussed in detail in the following sections.

Errors in plant species assemblage predictions vary in space

An important result is the decrease of the confidence intervals of the SDMs with increasing elevation. Individual SDMs are usually evaluated with metrics based on all of the predictions and locations (global accuracy as calculated with the AUC). However, our results strongly suggest that modelled plant species responses can show important variation in their accuracy when used for prediction across space. Previous findings on non-stationary species–environment relationships (Osborne *et al.*, 2007) suggest that this variation occurs for many taxa and ecosystems at different scales. Therefore, proper evaluations of SDM predictions should provide estimates of the variation in their quality over space in addition to global metrics (Hanspach *et al.*, 2011).

Our study further revealed varying abilities of stacked-SDMs (S-SDMs) to reproduce plant community patterns. We observed

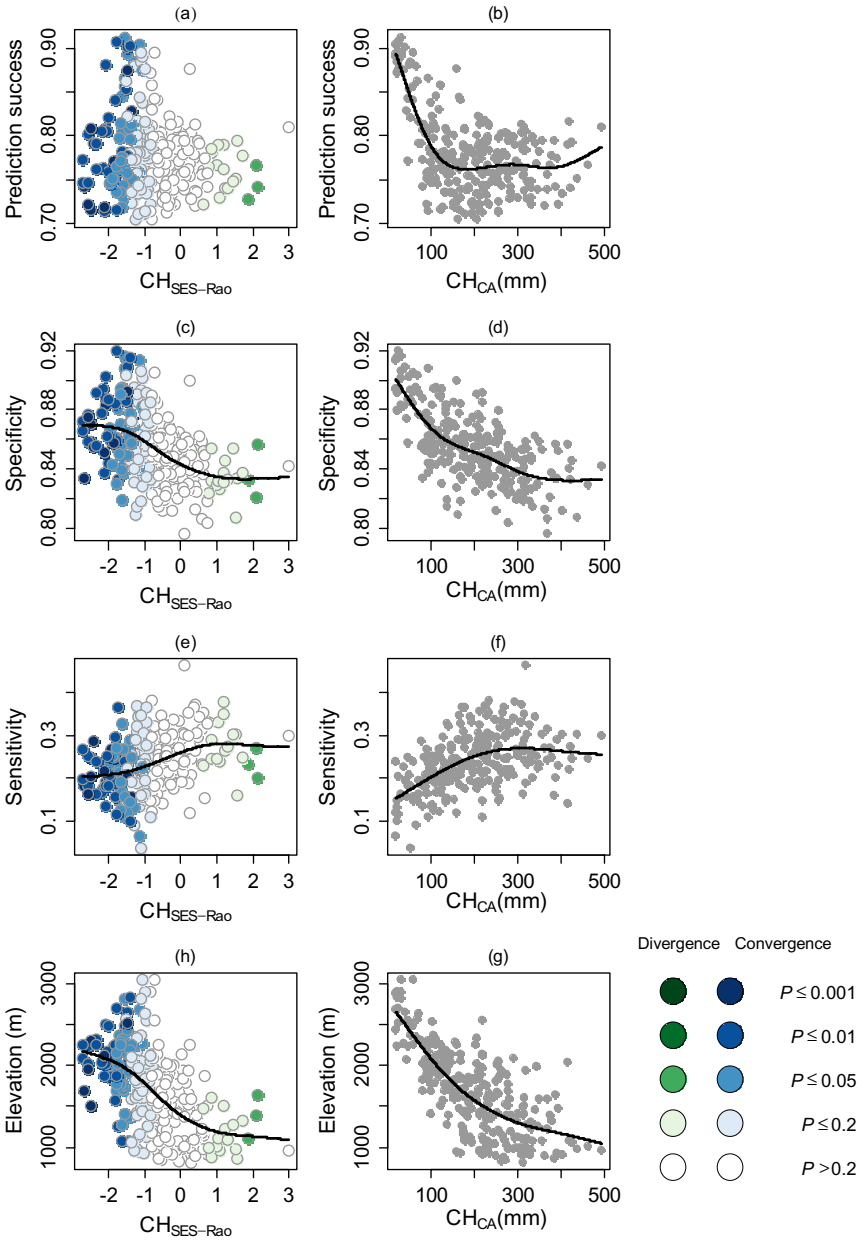


Figure 5 Variation in the accuracy of stacked-species distribution models (S-SDMs) with plant functional patterns in the evaluation plots. The plant functional trait being considered here is canopy height (CH, measured in mm). Its pattern was estimated using community aggregated mean (CH_{CA}, in mm) and evenness using the standardized effect size of the Rao index ($CH_{SES\text{-}Rao}$). Null distribution of trait values was estimated using 9999 simulations. Positive values of $CH_{SES\text{-}Rao}$ indicate that the observed distribution of canopy height tends to be more dissimilar than by chance (trait divergence indicative of limiting similarity) while negative values indicate more similarity (trait convergence indicative of environmental filtering). Evaluation metrics are the means over 9999 community samples based on the predicted probabilities of the presence of the 211 species. Solid lines represent significant trends fitted with generalized additive models.

an overall tendency of the S-SDMs to overpredict species richness, which is consistent with previous results based on different scales, regions and taxa (Feria & Peterson, 2002; Pineda & Lobo, 2009; Trotta-Moreu & Lobo, 2010) and is in line with theoretical expectations (Guisan & Rahbek, 2011). However, the novel finding here is that such overprediction can vary in space along environmental gradients. Indeed, the S-SDM approach did not reproduce the humpback curve of species richness along elevation, which is a common pattern when a large range of elevations is considered (Nogués-Bravo *et al.*, 2008). Instead, the S-SDMs predicted constant and high species richness from low to mid-range elevations and then a decrease towards high elevations. This prediction is consistent with better predictions of species richness reported by Guisan & Theurillat (2000) and by Dubuis *et al.* (2011) at high elevations. However, such a decrease was much slower than the observed one. For instance, the S-SDMs predicted a minimum number of species of approximately 18 at the highest elevations, while we recorded three to four species in the field. This mismatch resulted in an overprediction of species richness of up to 500%, which is translated into two opposite tendencies with regard to compositional evaluation metrics. On the one hand, the indices accounting for the correct predictions of absences (overall prediction success, kappa and specificity) showed better S-SDM performance towards higher elevations, as the S-SDMs predicted a decrease in species richness towards high elevations (although too slowly) but not towards low elevations. These findings suggest that the S-SDMs probably captured important ecological filters that restrict the composition of local species pools from the regional pool (Keddy, 1992; Zobel, 1997) at high elevations but likely not

at low elevations. On the other hand, the evaluation indices focusing on correct predictions of presence (sensitivity and the Jaccard index) reported worse prediction accuracy at high elevations than at mid-range and low elevations. This finding implies that strong limitations seem to affect S-SDMs towards high elevations, which need to be addressed.

Overall, considering all of the evaluation criteria, our results did not unequivocally verify our H1 hypothesis that the S-SDM performance improved towards high elevations. However, our results did not refute the reasoning behind this hypothesis. Rather, the highlighted patterns depict a more complex picture, as we discuss below.

Insufficient species filtering towards high elevations

The best assemblage specificity and prediction success were almost exclusively associated with strong environmental filtering (i.e. trait convergence) of small plant species towards unproductive environments at high elevations. There, the harsh climate is characterized by extreme temperatures and short growth periods. These conditions impose strong constraints on the physiology of plants, allowing only species of small stature to locally persist and coexist (Körner, 2003). Because climate and topography also influence community- and ecosystem-level processes, climatically stressful conditions may reduce the potential for competitive interactions to occur and the associated patterns of limiting similarity (i.e. trait divergence) for growth traits (Mitchell *et al.*, 2009). Our results showed that the strong climatic filtering of species could be reproduced using the S-SDMs, providing some support to the H2A hypothesis that the most accurate S-SDM predictions are predominantly associated with strong environmental filtering in climatically harsh environments. However, such support holds when considering the assemblage specificity and prediction success and when focusing on canopy height. When considering the assemblage sensitivity and other traits (specific leaf area and leaf dry matter content), the link between the S-SDM performance and particular assembly rules becomes hardly interpretable or non-existent. Moreover, the results tend to indicate that topo-climatic factors are not sufficient to restrict local species assemblages at very low values of species richness, as observed *in situ*. This finding suggests that important additional filters were not accounted for, even indirectly, in our set of predictors. For instance, a wide range of snow duration and snow depth can be found for similar climatic conditions because of snowdrift and its control by wind and topography (Litaor *et al.*, 2008). Geomorphological processes and pedogenesis are very good candidate predictors (Randin *et al.*, 2009) that would need to be added. Alternatively, the SDMs may have been flawed by data accuracy (Hanspach *et al.*, 2011). For instance, the computation and resolution of the climatic data cannot account for wind chill or small-scale microclimatic refuges (Scherrer & Körner, 2011). The prediction error at high elevations may also result from the decoupling of low-stature plants from the atmospheric conditions (Körner, 2003) captured by standard 2 m air-temperature weather stations. Consequently, the modelled species responses to temperature may probably not reproduce what species actually experience. In addition, in the alpine zone, vegetation cover is often scarce and patchy in a matrix of bare soil and rocks, implying that the evaluation of SDMs may be affected by a sampling effect at high elevations. Indeed, SDMs predict the potential presence of all species that can face local climatic conditions. Therefore, mismatches between observed and predicted assemblages may also result from the fact that vegetation records, conducted in relatively small sampling units (2 m × 2 m) at a given time, can easily miss species from the local pool that were (for some reason, such as local extinction by disturbances or too small plot size to capture all species) not observed in the sampled unit but were occurring nearby in the same habitat type (and thus were probably captured in other plots with similar conditions). Such sampling effects (in space and time) may prove especially important in sparse and disturbed vegetation, as found at the highest elevations, because growth is very slow there and thus recolonization after extinction is likely to be significantly delayed. Finer-resolution environmental predictors, and especially those related to substrate, would also be needed to better match the fine vegetation sampling units used in this study. Finally, we cannot exclude the hypothesis that a small part of the mismatch may also be due to overlooked species, i.e. species that were simply missed in the plots, because this type of error seems hardly avoidable in vegetation sampling, even by the best botanists (Vittoz *et al.*, 2010). However, this factor should remain constant along the elevation gradient, or should rather be problematic at lower elevations, where denser vegetation may hide some seedlings or small-stature plants. Therefore, it cannot solely account for the strong overprediction at higher elevations.

Non-equilibrium with climate towards low elevations

The reported links of S-SDM performance and community assembly rules also provided some support to our hypothesis H2B that less accurate predictions of S-SDMs are associated with a larger spectrum of assembly rules in the productive habitats found at low elevations. Indeed, we reported a much larger variety of assembly rules as the specificity, prediction success and elevation decreased. Moreover, patterns of trait convergence/divergence along elevation indicated that biotic assembly rules (i.e. a limited similarity of canopy height values) worked mainly in lowlands and that null assembly was more frequent in lowlands than in highlands. In mild climatic conditions, some places may be fertilized by agriculture, enhancing competitive interactions (Keddy *et al.*, 1997), leading either to a limited similarity of coexisting species or to apparent filtering of high-stature plants due to the exclusion of small species. In other locations, grazing may impose strong trait filtering (Díaz *et al.*, 2007) or release niche processes (Mason *et al.*, 2011) so that stochastic processes prevail. In addition, it is recognized that traditional practices sustain low levels of soil resources (Maurer *et al.*, 2006), resulting in strong environmental filtering due to harsh abiotic conditions in the lowlands, although these habitats often show high diversity. Together, these drivers of community

assembly can occur in different locations at the same time under the same mild climatic conditions and may thus overcome the influence of climate. In such situations, the species and community distributions are most likely not at equilibrium with the climate (at those elevations, forest is the natural climax in this region). S-SDMs based on topo-climatic factors alone exhibit greater uncertainty (Fig. 3) and necessarily lose their capacity to reliably filter the species from the landscape pool. In turn, species overprediction may have inflated the probability of predicting the correct species, resulting in greater assemblage sensitivity towards lower elevations.

Perspectives for biodiversity modelling

The main lesson emerging from these results is that the accuracy of species assemblage patterns reconstructed from topo-climatic SDMs can vary significantly in geographical space, especially where steep environmental gradients prevail, such as in mountainous landscapes. This finding strongly calls for a systematic evaluation of assemblage predictions even when individual SDMs show overall good predictive power. Moreover, we found that the increase or decrease of S-SDM prediction accuracy with elevation strongly depends on the evaluation criteria considered. While it is common practice to evaluate model predictions using a single general metric (e.g. kappa), our results clearly suggest that assemblage specificity and assemblage sensitivity (and their spatial variation) should also be assessed.

A second lesson learned from our findings is that S-SDMs based on topo-climatic factors may not represent species distribution and assembly drivers equally well along ecological gradients. For instance, in mild climatic conditions, such as those found in our study at lower elevations, species distributions and assemblies are in strong disequilibrium with the climate. Instead, other factors mediated by human activities may play a more prominent role. This disequilibrium is likely to be the case for other mountain ranges in the world (Nogués-Bravo *et al*., 2008). At higher elevations, species are likely to be under strong climatic determinism, but additional environmental filters and high-resolution data as well as more proximal climatic data may need to be considered.

As an important perspective, the generality of our findings should be tested by applying S-SDMs to other datasets along distinct altitudinal and latitudinal gradients. Further improvements of species assemblage predictions in a changing climate should account for additional local assembly drivers, such as competitive interactions, and additional environmental filtering other than climate and stochastic events (Guisan & Rahbek, 2011). Finally, new modelling approaches should be developed to incorporate the variability of the strength of assembly drivers over geographical and environmental spaces.

ACKNOWLEDGEMENTS

This research was conducted in the framework of the ECOCHANGE project funded by the Sixth European Framework Programme (grant GOCE-CT-2007–036866) and the BIOASSEMBLE project, funded by the Swiss National Science Foundation (SNF grant no. 31003A-125145 to A.G.). Large computations were performed at the Vital-IT (http://www.vital-it.ch) Centre for high-performance computing at the Swiss Institute of Bioinformatics. We are grateful to all the people involved in the data collection. We would like to thank D. D. Ackerly, R. Field, D. Currie and three anonymous referees for very helpful comments.

REFERENCES

Agrawal, A.A., Ackerly, D.D., Adler, F., Arnold, A.E., Cáceres, C., Doak, D.F., Post, E., Hudson, P.J., Maron, J., Mooney, K.A., Power, M., Schemske, D., Stachowicz, J., Strauss, S., Turner, M.G. & Werner, E. (2007) Filling key gaps in population and community ecology. *Frontiers in Ecology and the Environment*, **5**, 145–152.

Algar, A.C., Kharouba, H.M., Young, E.R. & Kerr, J.T. (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography*, **32**, 22–33.

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

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

Araújo, M.B. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. *Ecography*, **28**, 693–695.

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

de Bello, F., Lavergne, S., Meynard, C.N., Lepš, J. & Thuiller, W. (2010) The partitioning of diversity?: showing Theseus a way out of the labyrinth. *Journal of Vegetation Science*, **21**, 992–1000.

Callaway, R.M., Brooker, R.W., Choler, P., Kikvidze, Z., Lortie, C.J., Michalet, R., Paolini, L., Pugnaire, F.I., Newingham, B., Aschehoug, E.T., Armas, C., Kikodze, D. & Cook, B.J. (2002) Positive interactions among alpine plants increase with stress: a global experiment. *Nature*, **417**, 844–848.

Cornelissen, J.H.C., Lavorel, S., Garnier, E., Díaz, S., Buchmann, N., Gurvich, D.E., Reich, P.B., ter Steege, H., Morgan, H.D., van der Heijden, M.G.A., Pausas, J.G. & Poorter, H. (2003) Handbook of protocols for standardised and easy measurements of plant functional traits worldwide. *Australian Journal of Botany*, **51**, 335–380.

Díaz, S., Lavorel, S., McIntyre, S., Falczuk, V., Casanoves, F., Milchunas, D.G., Skarpe, C., Rusch, G., Sternberg, M., Noy-Meir, I., Landsberg, J., Zhang, W., Clark, H. & Campbell, B.D. (2007) Plant trait responses to grazing – a global synthesis. *Global Change Biology*, **13**, 313–341.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J. & Guisan, A. (2011) Predicting spatial patterns of plant species

richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122–1131.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

Engler, R., Randin, C.F., Thuiller, W. *et al.* (2011) 21st Century climate change threatens mountain flora unequally across Europe. *Global Change Biology*, **17**, 2330–2341.

Feria, A.T.P. & Peterson, A.T. (2002) Prediction of bird community composition based on point-occurrence data and inferential algorithms: a valuable tool in biodiversity assessments. *Diversity and Distributions*, **8**, 49–56.

Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.

Garnier, E., Cortez, J., Billès, G., Navas, M.-L., Roumet, C., Debussche, M., Laurent, G., Blanchard, A., Aubry, D., Bellmann, A., Neill, C. & Toussaint, J.-P. (2004) Plant functional markers capture ecosystem properties during secondary succession. *Ecology*, **85**, 2630–2637.

Gotelli, N.J. & McCabe, D.J. (2002) Species co-occurrence: a meta-analysis of J. M. Diamond'S assembly rules model. *Ecology*, **83**, 2091–2096.

Grime, J.P. (2001) *Plant strategies, vegetation processes and ecosystem properties*, 2nd edn. Wiley and Sons, New York.

Guisan, A. & Rahbek, C. (2011) SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.

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

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

Hanspach, J., Kühn, I., Schweiger, O., Pompe, S. & Klotz, S. (2011) Geographical patterns in prediction errors of species distribution models. *Global Ecology and Biogeography*, **20**, 779–788.

Hirzel, A. & Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331–341.

Keddy, P. (1992) Assembly and response rules: two goals for predictive community ecology. *Journal of Vegetation Science*, **3**, 157–164.

Keddy, P., Twolan-Strutt, L. & Shipley, B. (1997) Experimental evidence that interspecific competitive asymmetry increases with soil productivity. *Oikos*, **80**, 253–256.

Körner, C. (2003) *Alpine plant life – functional plant ecology of high mountain ecosystems*, 2nd edn. Springer, Heidelberg.

Litaor, M.I., Williams, M. & Seastedt, T.R. (2008) Topographic controls on snow distribution, soil moisture, and species diversity of herbaceous alpine vegetation, Niwot Ridge, Colorado. *Journal of Geophysical Research*, **113**, 1–10.

Maiorano, L., Falcucci, A., Zimmermann, N.E., Psomas, A., Pottier, J., Baisero, D., Rondinini, C., Guisan, A. & Boitani, L. (2011) The future of terrestrial mammals in the Mediterranean basin under climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2681–2692.

Mason, N.W.H., de Bello, F., Doležal, J. & Lepš, J. (2011) Niche overlap reveals the effects of competition, disturbance and contrasting assembly processes in experimental grassland communities. *Journal of Ecology*, **99**, 788–796.

Maurer, K., Weyand, A., Fischer, M. & Stöcklin, J. (2006) Old cultural traditions, in addition to land use and topography, are shaping plant diversity of grasslands in the Alps. *Biological Conservation*, **130**, 438–446.

Michalet, R., Brooker, R.W., Cavieres, L.A., Kikvidze, Z., Lortie, C.J., Pugnaire, F.I., Valiente-Banuet, A. & Callaway, R.M. (2006) Do biotic interactions shape both sides of the humped-back model of species richness in plant communities? *Ecology Letters*, **9**, 767–773.

Mitchell, M.G.E., Cahill, J.F. & Hik, D.S. (2009) Plant interactions are unimportant in a subarctic-alpine plant community. *Ecology*, **90**, 2360–2367.

Newbold, T., Gilbert, F., Zalat, S., El-Gabbas, A. & Reader, T. (2009) Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography*, **36**, 2085–2095.

Nogués-Bravo, D., Araújo, M.B., Romdal, T. & Rahbek, C. (2008) Scale effects and human impact on the elevational species richness gradients. *Nature*, **453**, 216–219.

Osborne, P.E., Foody, G.M. & Suárez-Seoane, S. (2007) Non-stationarity and local approaches to modelling the distributions of wildlife. *Diversity and Distributions*, **13**, 313–323.

Pakeman, R.J. & Quested, H.M. (2007) Sampling plant functional traits: what proportion of the species need to be measured? *Applied Vegetation Science*, **10**, 91–96.

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

R Development Core Team (2011) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Randin, C.F., Vuissoz, G., Liston, G.E., Vittoz, P. & Guisan, A. (2009) Introduction of snow and geomorphic disturbance variables into predictive models of alpine plant distribution in the western Swiss Alps. *Arctic Antarctic and Alpine Research*, **41**, 347–361.

Randin, C.F., Engler, R., Pearman, P.B., Vittoz, P. & Guisan, A. (2010) Using georeferenced databases to assess the effect of climate change on alpine plant species and diversity. *Data mining for global trends in mountain biodiversity* (ed. by E.M. Spehn and C. Körner), pp. 149–163. CRC Press, Boca Raton, FL.

Scherrer, D. & Körner, C. (2011) Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *Journal of Biogeography*, **38**, 406–416.

Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.

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

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

Trotta-Moreu, N. & Lobo, J.M. (2010) Deriving the species richness distribution of Geotrupinae (Coleoptera: Scarabaeoidea) in Mexico from the overlap of individual model predictions. *Environmental Entomology*, **39**, 42–49.

Vittoz, P., Bayfield, N., Brooker, R.W., Elston, D.A., Duff, E.I., Theurillat, J. & Guisan, A. (2010) Reproducibility of species lists, visual cover estimates and frequency methods for recording high-mountain vegetation. *Journal of Vegetation Science*, **21**, 1035–1047.

Weiher, E., Van Der Werf, A., Thompson, K., Roderick, M., Garnier, E. & Eriksson, O. (1999) Challenging Theophrastus: a common core list of plant traits for functional ecology. *Journal of Vegetation Science*, **10**, 609–620.

Westoby, M., Falster, D.S., Moles, A.T., Vesk, P.A. & Wright, I.J. (2002) Plant ecological strategies: some leading dimensions of variation between species. *Annual Review of Ecology and Systematics*, **33**, 125–159.

Wood, S.N. (2006) On confidence intervals for generalized additive models based on penalized regression splines. *Australian and New Zealand Journal of Statistics*, **48**, 445–464.

Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S. & Blackard, J.A. (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology*, **44**, 1057–1067.

Zobel, M. (1997) The relative role of species pools in determining plant species richness: an alternative explanation of species coexistence? *Trends in Ecology and Evolution*, **12**, 266–269.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Assessing the independence between the calibration and the evaluation datasets.
Appendix S2 Comparison of evaluation metrics between three elevation classes.
Appendix S3 Additional numerical properties of the metrics used to evaluate species assemblage predictions.
Appendix S4 Detailed variation of evaluation metrics with functional trait patterns.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

The Spatial Ecology group at the University of Lausanne (http://www.unil.ch/ecospat; led by A.G.) is specialized in niche-based spatial modelling of species, diversity and community distributions, using empirical data, statistical models and more dynamic approaches. A strong focus is given to use models and their predictions to support conservation management, such as of model-based sampling of endangered species, the assessment of the response of biodiversity to climate and land-use changes, and evaluating the invasive potential of neophyte species.

Author contributions: J.P. and A.G. designed the study. J.P., A.D., L.P., C.F.R., L.R., P.V. and A.G. collected the data. J.P. conducted the analyses with the help of A.D., and J.P. and A.G. wrote the manuscript with the help of A.D., L.P., L.M., C.F.R. and P.V.

Editor: Richard Field

ANNEX 3

Climate-based empirical models show biased predictions of butterfly communities along environmental gradients

Research

Ecography 35: 684–692, 2012
doi: 10.1111/j.1600-0587.2011.07047.x

Subject Editor: Miguel Araujo. Accepted 5 October 2011

Climate-based empirical models show biased predictions of butterfly communities along environmental gradients

Loïc Pellissier, Jean-Nicolas Pradervand, Julien Pottier, Anne Dubuis, Luigi Maiorano and Antoine Guisan

L. Pellissier (loic.pellissier@unil.ch), J.-N. Pradervand, J. Pottier, A. Dubuis, L. Maiorano and A. Guisan, Dept of Ecology and Evolution, Univ. of Lausanne, Bâtiment Biophore, CH-1015 Lausanne, Switzerland.

A better understanding of the factors that mould ecological community structure is required to accurately predict community composition and to anticipate threats to ecosystems due to global changes. We tested how well stacked climate-based species distribution models (S-SDMs) could predict butterfly communities in a mountain region. It has been suggested that climate is the main force driving butterfly distribution and community structure in mountain environments, and that, as a consequence, climate-based S-SDMs should yield unbiased predictions. In contrast to this expectation, at lower altitudes, climate-based S-SDMs overpredicted butterfly species richness at sites with low plant species richness and underpredicted species richness at sites with high plant species richness. According to two indices of composition accuracy, the Sorensen index and a matching coefficient considering both absences and presences, S-SDMs were more accurate in plant-rich grasslands. Butterflies display strong and often specialised trophic interactions with plants. At lower altitudes, where land use is more intense, considering climate alone without accounting for land use influences on grassland plant richness leads to erroneous predictions of butterfly presences and absences. In contrast, at higher altitudes, where climate is the main force filtering communities, there were fewer differences between observed and predicted butterfly richness. At high altitudes, even if stochastic processes decrease the accuracy of predictions of presence, climate-based S-SDMs are able to better filter out butterfly species that are unable to cope with severe climatic conditions, providing more accurate predictions of absences. Our results suggest that predictions should account for plants in disturbed habitats at lower altitudes but that stochastic processes and heterogeneity at high altitudes may limit prediction success of climate-based S-SDMs.

> *A collector who is a careful observer is often able to examine a terrain and to decide, intuitively as it were, whether a given butterfly will be found there … but this is a work of art rather than of science, and we would gladly know the components which make such predictions possible.*
>
> E. B. Ford (1945)

Understanding the factors that underlie the structure of natural communities is a prerequisite to adequately anticipating the threats of global change to ecosystems. Several statistical modelling approaches have been developed to explain and predict biodiversity patterns based on quantifiable species–environment relationships (reviewed by Ferrier and Guisan 2006). With the global availability of environmental maps, biodiversity models have been increasingly used to address conservation issues (Engler et al. 2011). Modelling at the community rather than the species level is often preferable because it remains relatively difficult to design proper biodiversity preservation measures based on a single indicator species (Ehrlich 1992). The derivation of reliable and practical conservation measures also requires that modelling studies be conducted at a sufficiently high resolution to provide accurate species–environment relationship information (Dennis and Hardy 1999, Pearson and Carroll 1999, Maiorano et al. 2011). However, relatively few studies have been performed at high resolution using ecologically reliable species and environmental datasets to evaluate species community predictions from statistical models.

Two approaches, ordination and species distribution modelling (SDM), have been proposed for predicting communities based on the occurrence of individual species (Ferrier and Guisan 2006). Ordination techniques allow the direct evaluation of community structures as they relate to species co-occurrences, corresponding to the so-called 'assemble first predict later' scheme by Ferrier and Guisan (2006). Similar techniques can be constrained by environmental variables, corresponding to the 'predict and assemble together' scheme in Ferrier and Guisan (2006). However, even the most advanced ordination techniques do not accurately fit patterns of species coexistence or correctly predict community compositions (Baselga and Araújo 2010). Alternatively, it is possible to model the responses of single species to their environment and then to stack these

individual models to predict communities, the 'predict first, assemble later' scheme by Ferrier and Guisan (2006).

This last approach is also called stacked species distribution modelling (S-SDM; Guisan and Rabhek 2011) and has been used to predict present and future distributions of species richness (Cumming 2000, Thuiller et al. 2005) and to estimate rates of species turnover under global climate change (Engler et al. 2011). Because S-SDM can predict not only species richness but also species composition, studies have also investigated whether this approach can reproduce community composition and/or species richness patterns (Pineda and Lobo 2009, Trotta-Moreu and Lobo 2010, Aranda and Lobo 2011). Yet, studies like this are still scarce, considering the relevance of such evaluation for the correct application of community models. S-SDM performance may be particularly well tested in a mountain study area that encompasses a wide variety of environmental conditions.

Butterflies represent an interesting group with which to examine the generality of S-SDM performance because butterflies are frequently used as indicator of biodiversity (Pearman et al. 2011). Climate plays a key role in determining patterns of butterfly distribution and assemblage (Stefanescu et al. 2004, 2011, Menéndez et al. 2007, Illián et al. 2010). In mountain environments, climate acts as a strong environmental filter, selecting for species that are able to cope with harsh conditions (Boggs and Murphy 1997). A recent study suggested that climatic variables dominate fine resolution butterfly distribution in mountain environments so that climatic SDM might be able to adequately predict butterfly distribution (Illián et al. 2010). Yet, the diversity and heterogeneity of both land cover type and land use type may also be important drivers of butterfly distribution and assemblage at fine resolution (Kerr et al. 2001, Bergman et al. 2004, White and Kerr 2006). Trophic dependency of butterfly species on a particular plant clade in highly coevolved and specialised systems may further affect species distribution and community structure (Ehrlich and Raven 1964). Butterflies that display specialised trophic interactions are typically more common in low-stress environments and where agricultural activities are not so intensive as to prevent plant-rich grasslands to develop (Kitahara and Fujii 1994). The impact of human activities on those ecosystems leads to a variety of disturbed habitats with impoverished plant communities (Stevens et al. 2004, Randin et al. 2009), potentially damaging the trophic relationship between butterflies and their host plants. Although a greater diversity of resources in plant-rich grasslands should promote a greater diversity of consumers (Hawkins and Porter 2003), lower plant species richness may restrict the butterfly community to the subset of its region that is predominantly constrained by climatic factors.

The ability of climate-based S-SDMs to predict butterfly communities is expected to depend on the relative importance of climate and other unaccounted-for variables, such as land use intensity, in conjunction with trophic interactions along environmental gradients. Evaluating the ability of climatic S-SDMs to predict butterfly communities may provide further insight into the drivers of butterfly distribution in mountain environments. If climate dominates butterfly distribution and community structure in mountain regions, the predictions from climatic S-SDMs should be unbiased. In contrast, if human activities influencing habitat structure and plant richness are also important components of butterfly distribution and community structure, predictions should be less accurate at low altitude sites where land use intensity limits the diversity of resources (i.e. grasslands plant species richness). In this study, we used a comprehensive dataset, collected with a robust sampling design, along an altitude gradient and considering state-of-the-art statistical methods, to evaluate the ability of climatic S-SDMs to predict butterfly communities in space, in particular their richness and composition, in a mountainous area of the western Swiss Alps.

Methods

Study area

The study area is located in the western Swiss Alps (Fig. 1). The altitude ranges from 1000 to 3210 m a.s.l. Along the altitudinal gradient, the vegetation belt sequence is typical of the calcareous Alps: a mountainous belt with mixed forests (*Fagus sylvatica*, European beech, and *Abies alba*, silver fir), a subalpine belt with coniferous forest (*Picea abies*, Norway spruce), an alpine belt with heath, meadow and grassland

Figure 1. Location of the study area in the western Alps of Switzerland. Black dots represent the butterfly sites that were sampled according to a random stratified design. The size of the dots is proportional to the butterfly species richness of the community.

vegetation (e.g. *Sesleria caerula*, *Nardus stricta*), and a nival belt with sparse vegetation cover dominated by taxa typical of high altitudes (e.g. *Saxifraga oppositifolia*). Vegetation in the study area has long been, and currently is, influenced by human land use. Pastures and meadows interspersed within forest patches are common in this region, from the valley bottoms up to the subalpine and lower alpine areas. Land use is particularly intense and diversified at lower altitudes, where open areas are intensively used as fertilised arable land; grasslands are dominated by *Arrhenatherum elatius* and *Poa* sp. At intermediate altitudes, agricultural activities are focused on cattle; fertilised pastures and meadows are dominated by *Cynosurus cristatus*, *Trisetum flavescens*, and *Heracleum sphondylium.* In areas providing limited economic revenues (e.g. steep slopes that are difficult to access using tractors), some species-rich grasslands dominated by *Bromus erectus* are still present, although the extent of these areas has been decreasing in the last three decades. Human land use is limited in areas near the treeline, and management is restricted to summer grazing in lower production pastures that are dominated by *Poa alpina*, *Sesleria caerulea*, or *Nardus stricta.*

Field sampling

We selected all of our sampling sites outside of forested areas, following a balanced stratified random sampling design based on altitude, slope and aspect (Hirzel and Guisan 2002). In 2009 and 2010, we sampled 192 sites from 1 June to 15 September during hours when butterflies are most active (10:00–17:00) and weather conditions were good (low wind, sunny, and high temperature, following Pollard and Yates 1993). Each site was visited every three weeks. For each 45-min visit, we walked through a square area of 50×50 m, collected butterflies with a net, and identified them in the field. Vegetation was exhaustively inventoried in a 4 m^2 subplot at the centre of each site (more details can be found in Dubuis et al. 2011). We inventoried a total of 131 butterfly species, including members of the Papilionoidea superfamily as well as Hesperiidae and Zygaenidae families. For the following analyses, we discarded species observed in fewer than 10 sites. We also removed any species displaying high mobility, as their presences were unlikely to be related to local environmental conditions (Pöyry et al. 2008), i.e. *Aglais urticae*, *Vanessa cardui*, *Vanessa atalanta*, *Pieris napi* and *Pieris brassicae.* A total of 77 species was retained for further analyses, and these species were considered the regional observed species pool (Supplementary material Appendix 1, Table A1).

Environmental variables

To model the species distribution, we selected a set of five climatic variables on the basis of butterfly ecology. Degree-days are related to a caterpillar's ability to emerge as an imago. The degree-days variable was calculated from monthly average temperatures following the method of Zimmermann and Kienast (1999). We also considered the minimum and maximum temperatures, which relates to physiological toleration of freezing and heating. The amount of solar radiation received has been suggested to be the most important descriptor of an area's butterfly distribution and richness (Turner et al. 1987, Hawkins 2003); solar radiation was calculated using the approach of Kumar et al. (1997). Finally, we included a moisture index, calculated following the method of Zimmermann and Kienast (1999). All climatic variables were calculated at 100 m resolution to fully encompass the 50×50 m butterfly sites. We verified that the correlation between variables was low to avoid multicollinearity problems. Additional details for the derivation of climatic predictors can be found in Zimmermann et al. (2007).

Community modelling and evaluation

We modelled the distribution of each species using four techniques implemented in the BIOMOD package (Thuiller et al. 2009) in R (R Development Core Team). We used the settings for each algorithm presented in Thuiller et al. (2009): a generalised linear model (GLM, McCullagh and Nelder 1989), a generalised additive model (GAM, Hastie and Tibshirani 1990), a gradient boosted model (GBM, Ridgeway 1999, Friedman et al. 2000) and a random forest model (RF, Breiman 2001). These techniques are all considered appropriate statistical methods for fitting presence-absence SDMs (see Elith et al. 2006 for an evaluation of GLM, GAM and GBM). We applied an ensemble forecasting framework by averaging all single model projections, weighted by AUC (Araújo and New 2007). Consensus methods based on average functions have been shown to provide a significant improvement in the accuracy of species distribution models, particularly when only the best methods are retained (Marmion et al. 2009). Independent evaluation is required to obtain a rigorous measure of SDM predictive power and S-SDM performance in predicting richness and community composition. In the absence of truly independent data, we ran a ten-fold cross-validation. We randomly split the dataset into ten partitions, successively calibrated the models using 90% of the data, and sequentially predicted the species and communities based on the remaining 10% of the data. We calculated the SDMs predictive power for each species as the area under the curve (AUC) of a receiver operating characteristic plot (ROC; Swets 1988).

To obtain predictions of species distributions as binary presence or absence independent of a threshold, we used the rbinom function in R (R Development Core Team) to draw presence or absence from the probabilities provided by the SDMs in the binomial distribution. This solution was proposed by Dubuis et al. (2011) and was shown to reduce overprediction of species richness. If a species is predicted to occur with a probability of 0.7, we use a binomial distribution (i.e. the 'rbinom' function in R) to draw 0 or 1 with a distribution of probabilities with $p = 0.7$ (i.e. same process as flipping a coin, but here the probabilities of the two outcomes are different). The choice of a threshold is still highly debated in the literature and an approach that does not need any threshold and use directly the probabilities of occurrences produced by SDMs would be seen as more objective. Using this method, we obtained a community matrix for each draw with presence-absence records that could be

used to evaluate S-SDM performance. By doing this for all species, this will predict one potential realization of the community given the probability of occurrence of each species. By repeating this process many times, we obtain a large number of possible realizations of the potential community at each location. This binomial resampling approach allows simulating the different endpoints of community assembly advocated by Guisan and Rahbek (2011). We repeated this process 1000 times and computed the average to account for the variability in the random draws.

To evaluate the ability of S-SDM to predict butterfly species richness, we correlated the observed species richness and the average predicted species richness calculated from the 1000 draws using a linear model (LM). We also correlated the residuals between observed and predicted butterfly species richness to altitude and plant species richness using a LM and tested the effects with F-tests. We included an interaction term between the two explanatory variables because we did not expect plant species richness to have an identical effect at high altitudes and low altitudes. Finally, we evaluated the ability of S-SDM to predict butterfly communities based on variations in altitude and plant species richness. We computed two metrics describing the ability of S-SDMs to predict community composition by comparing observed and predicted communities for each site as follows:

$$\text{Prediction success} = \frac{a+d}{N} \quad (1)$$

$$\text{Sorensen index} = \frac{2a}{2a+b+c} \quad (2)$$

where a is the number of species that are both observed and predicted as being present, b is the number of species observed as present but predicted as absent, c is the number of species observed as absent but predicted as present, d is the number of species that are both absent and predicted as absent, and N is the total number of species.

Prediction success evaluates both the ability to filter out absent species and predict the occurrences of observed species from the entire regional species pool. By considering both presences and absences, this index also accounts for the ability of S-SDMs to predict species richness. In contrast, the Sorensen index considers only the ability to correctly predict species presences. To assess whether S-SDM prediction of butterfly communities was biased, we correlated S-SDM prediction success and Sorensen index averaged across 1000 draws to altitude and plant species richness using a LM coupled with an F-test, as detailed above.

Results

Species distribution models

Overall, the climatic variables included in the models were able to explain the distribution of most species. The predictive capacities of the SDMs were not identical across all modelling techniques, and the predictive abilities were ranked from poor to very good (Swets 1988), with AUCs ranging from low (0.50) to high (0.95) and an average across the 77 species of 0.75 (GAM: 0.76, GBM: 0.75, GLM: 0.76, RF: 0.74; Supplementary material Appendix 1, Table A1). Only a few species had a very poor AUC, and including them or not in the following analyses did not influence the results.

Predictions of community richness and composition

Predicted species richness was significantly correlated with observed species richness, but the relationship was not the expected 1:1 relationship (predicted = 11.1 + 0.26 × observed, $R^2 = 0.56$, $p < 0.0001$). The residual values between observed and predicted butterfly species richness were significantly and negatively correlated with plant richness but only marginally significantly correlated with altitude (Table 1). The S-SDMs overpredicted butterfly species richness in plant-poor grasslands and underpredicted butterfly species richness in plant-rich grasslands. The residuals between observed and predicted butterfly species richness were highly dispersed around the expected zero value, and the dispersion diminished progressively with altitude (Fig. 2, 3). Overall, overprediction was more common than underprediction.

The prediction success index was significantly correlated with altitude and plant species richness (Table 1). According to this index, the best predictions were achieved at high altitude, but also in grasslands with higher plant species richness. The interaction term was also significant, indicating that prediction success was greater in higher altitude communities, which are characterised by low plant species richness, than in communities at lower altitudes. At lower altitudes, prediction success was greater in sites with higher plant species richness (Table 1, Fig. 4). The Sorensen index was also significantly correlated with altitude and plant species richness, but in contrast to the prediction success index, it achieved higher scores for communities at lower altitudes (Fig. 4). The difference between the two indices arose from whether or not the ability to predict absences in communities was considered; this difference increases as altitude increases (Fig. 5). Note that we also considered an approach based on

Table 1. Results of the F-tests on the linear relationships between the butterfly richness residuals, the prediction success and Sorensen index related to altitude and plant species richness respectively.

	Altitude			Plant richness			Interaction		
	Estimate	F	p	Estimate	F	p	Estimate	F	p
Richness residuals	6.77E-04	2.7	0.09	−1.08E-01	11	0.001	1.36E-05	0.02	0.87
Prediction success	1.27E-04	179.3	<0.0001	1.45E-03	30.6	<0.0001	−1.53E-06	7	0.008
Sorensen	−4.76E-05	10.5	0.001	1.63E-03	9.7	0.002	5.46E-08	0.001	0.97

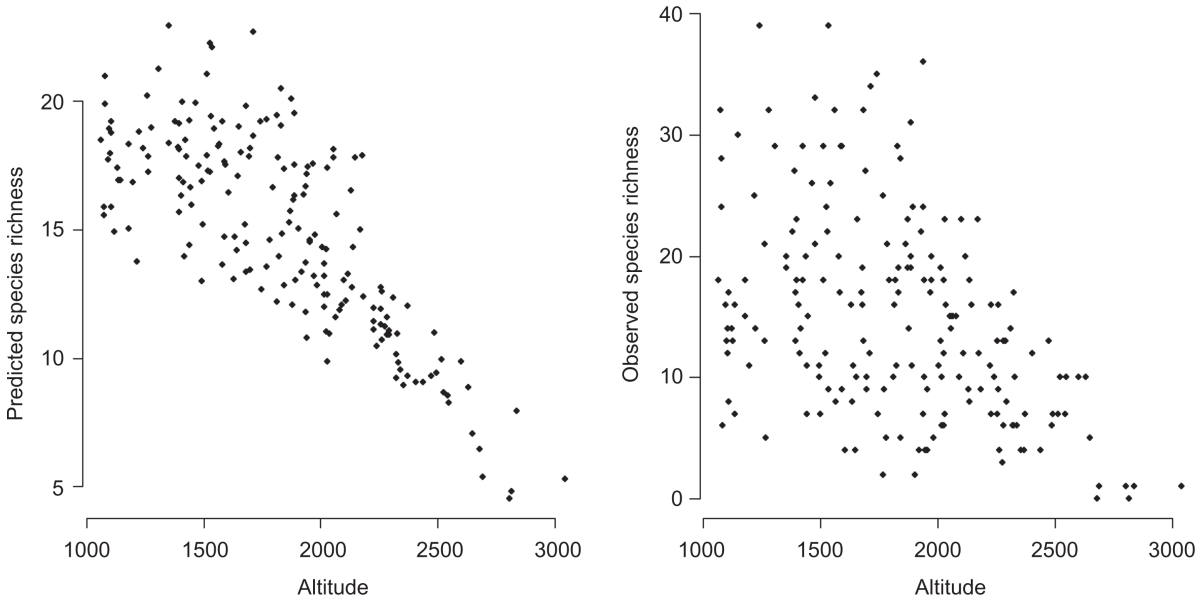


Figure 2. Variation of observed and predicted (averaged over 1000 community draws) species richness along an altitude gradient.

thresholds to transform model probabilities into presences and absences, obtaining comparable results (Supplementary material Appendix 1).

Discussion

Using state-of-the-art modelling techniques and considering only climate variables, we predicted the distribution and community structure of the complete pool of butterfly species in a mountain range in the western Swiss Alps. Our study revealed that S-SDMs gave biased predictions for community richness and composition along altitude and plant species richness gradients. Even though the compositions of several communities were predicted accurately, climate-based S-SDMs do not appear to be entirely able to predict butterfly communities, in mountain environments in terms of richness and composition.

The outcome of the composition analysis varied between the two metrics used. The prediction success index was measured as a standard matching coefficient between presences and absences and the Sorensen index, which considers only species presences. The Sorensen index indicated that S-SDMs tend to be more accurate at low altitude. The Sorensen index indicated that the models performed especially poorly at the highest altitudes, which are almost devoid of vegetation. In severe climates, the environment is less stable, both temporally and spatially, with patches of vegetation interspersed with bare rock and scree (Körner 2003). In turn, the butterfly communities in these areas

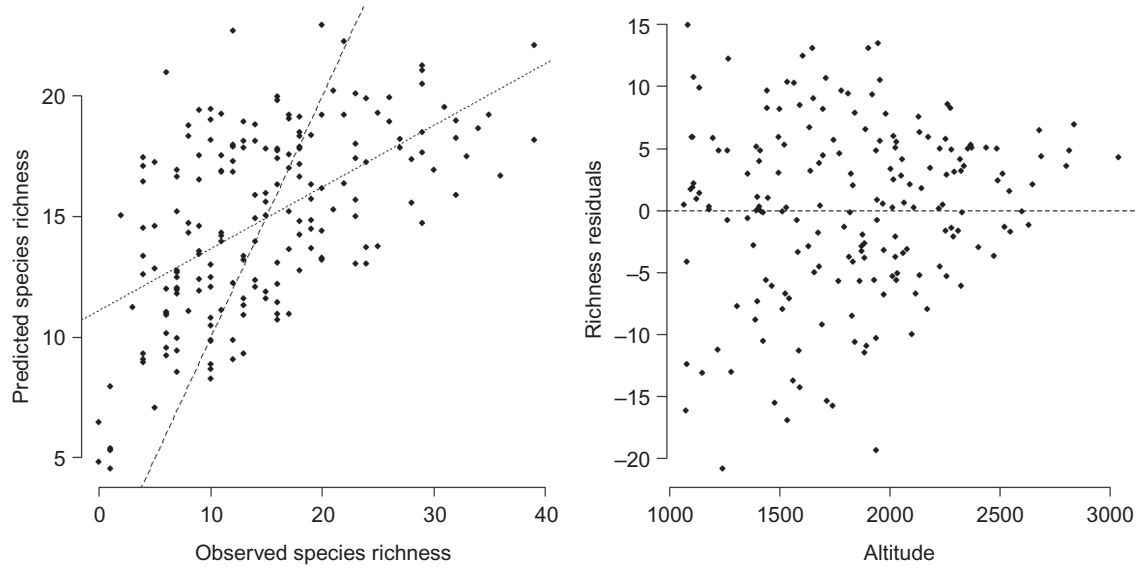


Figure 3. Relationship between the predicted (averaged over 1000 community draws) and observed butterfly species richness. The dotted line corresponds to the regression between observed and predicted species richness; dashed lines correspond to the 1:1 line. The right plot represents the residuals between observed and predicted species richness along altitude with the line at the expected zero value.

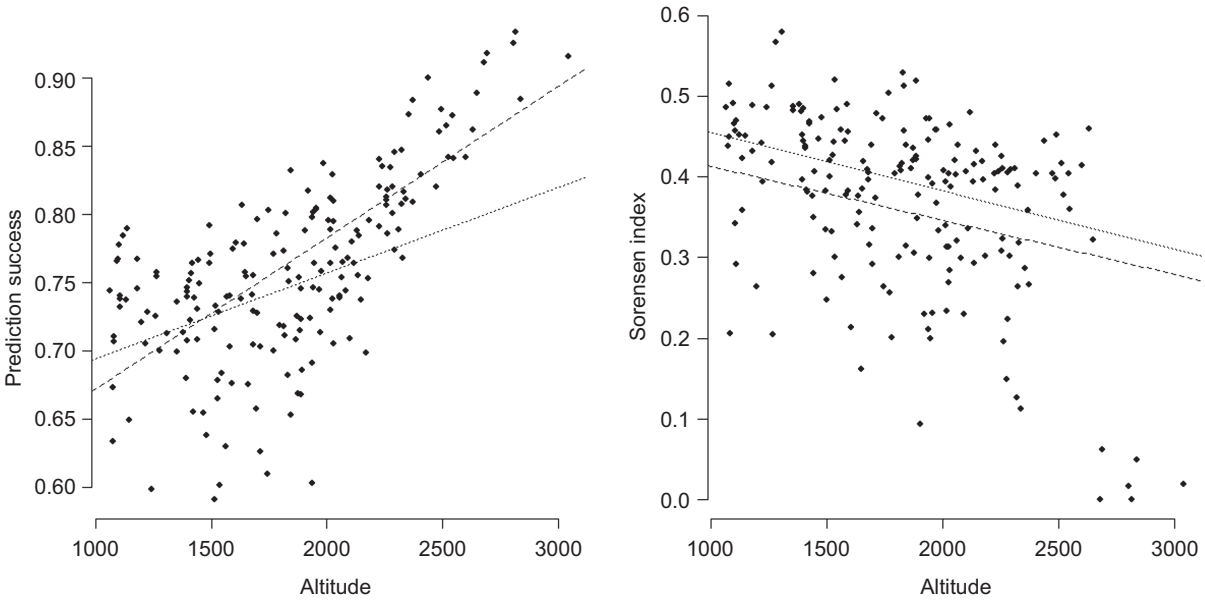


Figure 4. Prediction success (left) and Sorensen index (right), relative to altitude. The dotted and dashed lines correspond to the regression for sites with plant species richness under and over the average, respectively (n = 32 plant species).

may be more stochastic (Gutiérrez and Menéndez 1998), making community prediction more challenging. Indeed, when drawing the presences and absences from the model probabilities in higher altitude communities, the predicted communities were much more variable between draws (Supplementary material Appendix 1, Fig. A1). In contrast, the greater butterfly richness at lower altitude allows the S-SDM to more frequently predict the correct species composition from the entire species pool.

The Sorensen index was first proposed in community ecology to deal with the 'double-zero' problem that arises when comparing two observed communities (Legendre

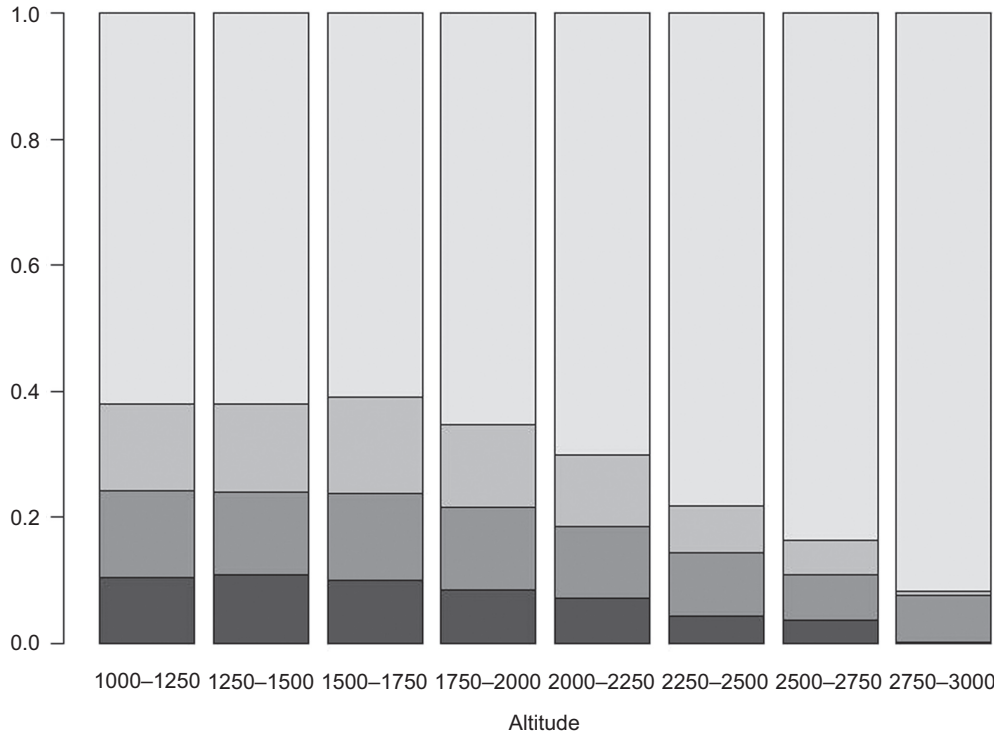


Figure 5. Bar plot of the four components of community prediction based on a contingency table. The bars, listed from bottom to top, represent the following species proportions: a, the proportion of species that were both observed and predicted as present; b, the proportion of species that were not observed but were predicted as present (over-prediction), c, the proportion of species that were observed but were predicted as absent (under-prediction); and d, the proportion of species that were both absent and predicted as absent. Each bar corresponds to a 250 m altitude interval.

and Legendre 1998). Although identical absences of species may be uninformative when comparing two observed communities, both the ability to predict correctly present species and the ability to filter out absent species are relevant when comparing a community predicted from a model to its observed counterpart. The prediction success metric considers both presence and absence predictions and showed that community composition was more accurately predicted at high altitudes. At high altitudes, the climate is the main factor that determines species presence or absence, depending on whether the species is able to cope with severe conditions or not. Only a limited number of species with greater tolerance to cold temperatures can survive under these conditions, and many species are thus filtered out from the regional species pool (Boggs and Murphy 1997). In these cases, climatic variables are able to represent the physiological limitations of butterflies in SDMs, and S-SDMs can therefore predict community compositions, particularly species absences, with higher accuracy. Nevertheless, overprediction remains several times larger than the proportion of correctly predicted presences. Climate is not the only environmental filter and extended rock cover may favour species tolerating rocky environment (e.g. *Erebia pluto*, *E. gorge*) and exclude species requiring dense vegetation cover. As a consequence, climatic S-SDMs may represent a good filter to predict high altitude fauna, but is not totally sufficient either. Also, at low altitude, the larger pool of butterfly species causes a higher number of possible combinations, which may impede a good success of predictions of communities. Moreover, a drawback of the prediction success index is its dependence on the regional species pool considered. One can arbitrarily increase the regional species pool including more absent species and subsequently increasing the predictive success. As a consequence, the regional species pool should be carefully identified by considering the pool of species that could have dispersed to a given region (Guisan and Rahbek 2011).

At lower altitude, both indices exhibited greater prediction accuracy in plant-rich grasslands. MacArthur (1972) hypothesised that while climatic factors are more important determinants of species range limits in harsher environmental conditions, biotic interactions are more likely to control distribution limits under less stressful conditions. For example, Merrill et al. (2008) found that the distribution of the butterfly *Aporia crataegi* in a mountainous area in Spain corresponded with that of its larval host plant in the less climatically stressful part of the gradient, but that the species was excluded from areas with dry conditions despite the presence of its host plant. In our study area, climate is a limiting factor at high altitude, where cold and moist conditions act as strong physiological stresses. In contrast, specialised trophic interactions are likely to be more important in the less stressful environments found at lower altitudes. Consequently, climatic S-SDMs may miss information that is important for predicting butterfly community composition at lower altitudes. In particular, resource limitation can reduce the accuracy of climatic S-SDMs in plant-poor grasslands.

Considering that greater plant diversity at a given site is expected to support a greater diversity of specialist butterflies (Hawkins and Porter 2003), only a subset of the species predicted by S-SDMs could occur in areas with lower plant species richness. Consistent with this, our results showed species overprediction in grasslands with poorer plant richness. McDonnell et al. (1993) predicted that trophic interactions are the most significant limiting factor for caterpillars in rural areas relative to more physically stressed environments. This difference is partially due to land use intensity that leads to poor and homogeneous plant communities (Stevens et al. 2004) and therefore to a decrease in host plant species richness. Intense land-use patterns prevent butterfly species that exploit specialised resources from occupying the habitat (Kitahara and Fujii 1994). Thus, human disturbances may have strong impacts on trophic interactions that influence the distributions of specialist butterflies (Bergman et al. 2004, Ekroos et al. 2010). Because trophic interactions are expected to drive species occupancy in climatically unconstrained habitats (Merrill et al. 2008), erroneous predictions of butterfly community composition and over-prediction of species richness may arise at site with intense land use. In plant-rich grasslands, overprediction is reduced, allowing increased prediction accuracy. At a larger spatial scale, Hanspach et al. (2011) showed that prediction errors in plant species distribution over Europe was concentrated in areas with intense land use.

S-SDMs also underpredicted species richness in several grasslands with particularly high plant species richness. Due to heterogeneity in land use intensity, variations in both plant and butterfly species richness can be high at lower altitudes. Climate-based SDM is not able to account for much of this variation between communities. Considering that the grasslands in the study area at lower altitudes may vary from intense land use to more relaxed land use, climate-based SDMs tend to produce average values for the presences and absences in communities under similar climatic conditions but different plant species richness. As a consequence, underprediction may also arise when presences and absences are randomly drawn from the less contrasted model probabilities.

In conclusion, we have shown that climate based S-SDMs are not able to accurately predict butterfly species richness and composition in mountain environments that are under human influence. The predictions showed important biases along gradients of altitude and plant species richness, and our results suggest that predictions should account for trophic resources in addition to climate factors. Hanspach et al. (2011) suggested that the standard measure of predictive performance per species should be accompanied by a complementary analysis of spatial bias (Rocchini et al. 2011). Our results concur, indicating that evaluations of the predictive performance of community models along environmental gradients might provide important insights into the drivers of community assembly, such as land use intensity or biotic interactions (Araújo and Luoto 2007, Schweiger et al. 2008, Pellissier et al. 2010). As changes in land use are likely to continue to accelerate and intensify in Europe (Van Swaay et al. 2006), unfertilised semi-natural grasslands are expected to become increasingly marginal (Falcucci et al. 2007). One possible application of an S-SDM approach could be to select potential sites for agri-environmental schemes (AES) that are designed to optimise the positive effects of organic farming at the landscape scale (Rundlöf et al. 2008). Because the effectiveness of these schemes

depends on the landscape context (Holzschuh et al. 2008), the identification of areas suitable for AES could benefit from a priori biodiversity information derived by S-SDMs. More generally, S-SDM community models can be combined with land use change models or projected for the future according to land use and climate change scenarios to forecast community responses to global change.

Acknowledgements – We are grateful to Alex Von Ungern-Steinberg, Valéry Uldry and Virginie Favre as well as all the people who contributed in the fieldwork. We also thank the referees for helpful comments on the manuscript. This study was supported by the European Commission (ECOCHANGE project, contract no. FP6 2006 GOCE 036866) and NSF grant no. 31003A-125145 (BIO-ASSEMBLE project).

References

Aranda, S. C. and Lobo, J. M. 2011. How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. – Ecography 34: 31–38.

Araújo, M. B. and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. – Global Ecol. Biogeogr. 16: 743–753.

Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – Trends Ecol. Evol. 22: 42–47.

Baselga, A. and Araújo, M. B. 2010. Do community-level models describe community variation effectively? – J. Biogeogr. 37: 1842–1850.

Bergman, K. O. et al. 2004. Landscape effects on butterfly assemblages in an agricultural region. – Ecography 27: 619–628.

Boggs, C. L. and Murphy, D. D. 1997. Community composition in mountain ecosystems: climatic determinants of montane butterfly distributions. – Global Ecol. Biogeogr. 6: 39–48.

Breiman, L. 2001. Random forests. – Machine Learn. 45: 5–32.

Cumming, G. S. 2000. Using habitat models to map diversity: pan-African species richness of ticks (Acari: Ixodida). – J. Biogeogr. 27: 425–440.

Dennis, R. L. H. and Hardy, P. B. 1999. Targeting squares for survey: predicting species richness and incidence of species for a butterfly atlas. – Global Ecol. Biogeogr. 8: 443–454.

Dubuis, A. et al. 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking approaches. – Divers. Distrib. 17: 1122–1131.

Ehrlich, P. R. 1992. Population biology of checkerspot butterflies and the preservation of global biodiversity. – Oikos 63: 6–12.

Ehrlich, P. R. and Raven, P. H. 1964. Butterflies and plants: a study in coevolution. – Evolution 18: 586–608.

Ekroos, J. et al. 2010. Homogenization of lepidopteran communities in intensively cultivated agricultural landscapes. – J. Appl. Ecol. 47: 459–467.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – Ecography 29: 129–151.

Engler, R. et al. 2011. Climate change impacts on European mountain plant diversity. – Global Change Biol. in press.

Falcucci, A. et al. 2007. Changes in land-use/land-cover patterns in Italy and their implications for biodiversity conservation. – Landscape Ecol. 22: 617–631.

Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. – J. Appl. Ecol. 43: 393–404.

Friedman, J. H. et al. 2000. Additive logistic regression: a statistical view of boosting. – Ann. Stat. 28: 337–374.

Guisan, A. and Rahbek, C. 2011. SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. – J. Biogeogr. in press.

Gutiérrez, D. and Menéndez, R. 1998. Stability of butterfly assemblages in relation to the level of numerical resolution and altitude. – Biol. Conserv. 7: 967–979.

Hanspach, J. et al. 2011. Geographical patterns in prediction errors of species distribution models. – Global Ecol. Biogeogr. in press.

Hastie, T. and Tibshirani, R. 1990. Generalized additive models. – Chapman and Hall.

Hawkins, B. A. 2003. Water–energy balance and the geographic pattern of species richness of western Palearctic butterflies. – Ecol. Entomol. 41: 678–686.

Hawkins, B. A. and Porter, E. E. 2003. Does herbivore diversity depend on plant diversity? The case of California butterflies. – Am. Nat. 161: 40–49.

Hirzel, A. and Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. – Ecol. Model. 157: 331–341.

Holzschuh, A. et al. 2008. Agricultural landscapes with organic crops support higher pollinator diversity. – Oikos 117: 354–361.

Illián, J. G. et al. 2010. The contributions of topoclimate and land cover to species distributions and abundance: fine-resolution tests for a mountain butterfly fauna. – Global Ecol. Biogeogr. 19: 159–173.

Kerr, J. T. et al. 2001. Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. – Proc. Natl Acad. Sci. USA 98: 11365–11370.

Kitahara, M. and Fujii, K. 1994. Biodiversity and community structure of temperate butterfly species within a gradient of human disturbance: an analysis based on the concept of generalist vs. specialist strategies. – Res. Popul. Ecol. 36: 187–199.

Körner, C. 2003. Alpine plant life. – Springer.

Kumar, L. et al. 1997. Modelling topographic variation in solar radiation in a GIS environment. – Int. J. Geogr. Inf. Sci. 11: 475–497.

Legendre, P. and Legendre, L. 1998. Numerical ecology. – Elsevier.

MacArthur, R. H. 1972. Geographical ecology. – Harper and Row.

Maiorano, L. et al. 2011. The future of terrestrial mammals in the Mediterranean basin under climate change. – Phil. Trans. Soc. B 27: 2681-2692.

Marmion, M. et al. 2009. Evaluation of consensus methods in predictive species distribution modelling. – Divers. Distrib. 15: 59–69.

McCullagh, P. and Nelder, J. A. 1989. Generalized linear models, 2nd ed. – Chapman and Hall.

McDonnell, M. J. et al. 1993. The application of the ecological gradient paradigm to the study of urban effects. – In: McDonnell, M. J. and Pickett, S. T. A. (eds), Humans as a components of ecosystems. Springer, pp. 175–189.

Menéndez, R. et al. 2007. Direct and indirect effect of climate and habitat factors on butterfly diversity. – Ecology 88: 605–611.

Merrill, R. M. et al. 2008. Combined effects of climate and biotic interactions on the elevational range of a phytophagous insect. – J. Anim. Ecol. 77: 145–155.

Pearman, P. B. et al. 2011. Impacts of climate change on Swiss biodiversity: an indicator taxa approach. – Biol. Conserv. 144: 866–875.

Pearson, D. L and Carroll, S. S. 1999. The influence of spatial scale on cross-taxon and prediction congruence patterns accuracy of species richness. – J. Biogeogr. 26: 1079–1090.

Pellissier, L. et al. 2010. Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. – Ecography 33: 1004–1014.

Pineda, E. and Lobo, J. M. 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. – J. Anim. Ecol. 78: 182–190.
Pollard, E. and Yates, T. J. 1993. Monitoring butterflies for ecology and conservation. – Chapman and Hall.
Pöyry, J. et al. 2009. Species traits explain recent range shifts of Finnish butterflies. – Global Change Biol. 15: 732–743.
Randin, C. F. et al. 2009. Land use improves spatial predictions of mountain plant abundance but not presence-absence. – J. Veg. Sci. 20: 996–1008.
Ridgeway, G. 1999. The state of boosting. – Comp. Sci. Stat. 31: 172–181.
Rocchini, D. et al. 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. – Prog. Phys. Geogr. 35: 211–225.
Rundlöf, M. et al. 2008. Local and landscape effects of organic farming on butterfly species richness and abundance. – J. Appl. Ecol. 45: 813–820.
Schweiger, O. et al. 2008. Climate change can cause spatial mismatch of trophically interacting species. – Ecology 89: 3472–3479.
Stefanescu, C. et al. 2004. Butterfly species richness in the north-west Mediterranean Basin: the role of natural and human-induced factors. – J. Biogeogr. 18: 259–915.
Stefanescu, C. et al. 2011. Determinants of species richness in generalist and specialist Mediterranean butterflies: the negative synergistic forces of climate and habitat change. – Ecography in press.

Supplementary material (Appendix E7047 at <www.oikosoffice.lu.se/appendix>). Appendix 1.

Stevens, C. J. et al. 2004. Impact of nitrogen deposition on the species richness of grasslands. – Science 303: 1876–1879.
Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. – Science 240: 1285–1293.
Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. – Proc. Natl Acad. Sci. USA 102: 8245–8250.
Thuiller, W. et al. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. – Ecography 32: 369–373.
Trotta-Moreu, N. and Lobo, J. M. 2010. Deriving the species richness distribution of Geotrupinae (Coleoptera: Scarabaeoidea) in Mexico from the overlap of individual model predictions. – Environ. Entomol. 39: 42–49.
Turner, J. R. G. et al. 1987. Does solar energy control organic diversity? Butterflies, moths and the British climate. – Oikos 48: 195–205.
Van Swaay, C. et al. 2006. Biotope use and trends of European butterflies. – J. Insect Conserv. 10: 189–209.
White, P. and Kerr, J. T. 2006. Contrasting spatial and temporal global change impacts on butterfly species richness during the 20th century. – Ecography 29: 908–918.
Zimmermann, N. E. and Kienast, F. 1999. Predictive mapping of alpine grasslands in Switzerland: species versus community approach. – J. Veg. Sci. 10: 469–482.
Zimmermann, N. E. et al. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. – J. Appl. Ecol. 44: 1057–1067.

ANNEX 4

Ecological assembly rules in plant communities - approaches, patterns and prospects

BIOLOGICAL REVIEWS
Cambridge Philosophical Society
Biol. Rev. (2012), **87**, pp. 111–127.
doi: 10.1111/j.1469-185X.2011.00187.x

Ecological assembly rules in plant communities—approaches, patterns and prospects

Lars Götzenberger[1,*], Francesco de Bello[2], Kari Anne Bråthen[3], John Davison[1], Anne Dubuis[4], Antoine Guisan[4,5], Jan Lepš[6,7], Regina Lindborg[8,9], Mari Moora[1], Meelis Pärtel[1], Loic Pellissier[4], Julien Pottier[4], Pascal Vittoz[4], Kristjan Zobel[1] and Martin Zobel[1]

[1] *Department of Botany, University of Tartu, Lai 40, 51005 Tartu, Estonia*
[2] *Institute of Botany, Czech Academy of Sciences, Dukelská 135, 379 82 Třeboň, Czech Republic*
[3] *Department of Arctic and Marine Biology, University of Tromsø, 9037 Tromsø, Norway*
[4] *Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland*
[5] *Institute of Geology & Paleontology, University of Lausanne, 1015 Lausanne, Switzerland*
[6] *Department of Botany, Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic*
[7] *Institute of Entomology, Biology Center, Academy of Sciences of the Czech Republic, Branišovská 31, 370 05 České Budějovice, Czech Republic*
[8] *Department of Physical Geography and Quaternary Geology, Stockholm University, SE-106 91 Stockholm, Sweden*
[9] *Stockholm Resilience Centre, Stockholm University, SE-106 91 Stockholm, Sweden*

ABSTRACT

Understanding how communities of living organisms assemble has been a central question in ecology since the early days of the discipline. Disentangling the different processes involved in community assembly is not only interesting in itself but also crucial for an understanding of how communities will behave under future environmental scenarios. The traditional concept of assembly rules reflects the notion that species do not co-occur randomly but are restricted in their co-occurrence by interspecific competition. This concept can be redefined in a more general framework where the co-occurrence of species is a product of chance, historical patterns of speciation and migration, dispersal, abiotic environmental factors, and biotic interactions, with none of these processes being mutually exclusive.

Here we present a survey and meta-analyses of 59 papers that compare observed patterns in plant communities with null models simulating random patterns of species assembly. According to the type of data under study and the different methods that are applied to detect community assembly, we distinguish four main types of approach in the published literature: species co-occurrence, niche limitation, guild proportionality and limiting similarity.

Results from our meta-analyses suggest that non-random co-occurrence of plant species is not a widespread phenomenon. However, whether this finding reflects the individualistic nature of plant communities or is caused by methodological shortcomings associated with the studies considered cannot be discerned from the available metadata.

We advocate that more thorough surveys be conducted using a set of standardized methods to test for the existence of assembly rules in data sets spanning larger biological and geographical scales than have been considered until now. We underpin this general advice with guidelines that should be considered in future assembly rules research. This will enable us to draw more accurate and general conclusions about the non-random aspect of assembly in plant communities.

Key words: community assembly, competition, co-occurrence, guild proportionality, limiting similarity, meta-analysis, niche limitation, null model, interspecific interaction.

* Address for correspondence (Tel.: +372 7376090; Fax: +372 7376222; E-mail: lars@ut.ee).

CONTENTS

I. SETTING THE SCENE: ASSEMBLY RULES

(1) Introduction

Since the early days of plant community ecology there have been contrasting views concerning the processes that assemble plant communities. Some early thinkers viewed plant communities as "complex organisms", reflecting the notion that communities are not assembled randomly and only exist in certain combinations (Clements, 1916; Phillips, 1931). By contrast, the vast spatial and temporal variability of plant community composition led Gleason (1926, p. 16) to propose precisely the opposite concept: that "... an association is not an organism... but merely a *coincidence*". Recent thinking places those two views at opposing sides of a conceptual gradient, with both stochastic and deterministic processes generally believed to contribute to the formation of community assemblages (Lortie *et al.*, 2004; Vergnon, Dulvy & Freckleton, 2009; Kembel, 2009).

In the 1980s and 1990s, a discussion about non-random processes underlying community assembly was initiated (Ricklefs, 1987; Keddy, 1992; Eriksson, 1993; Zobel, 1997; Diaz, Cabido & Casanoves, 1998; Belyea & Lancaster, 1999). This discussion is ongoing (Grime, 2006; Wilson, 2007; Ricklefs, 2008; Brooker *et al.*, 2009), and the search for principles to explain the assembly of local communities remains a major focus in plant community ecology (Hubbel, 2001; Fargione, Brown & Tilman, 2003; Bruun & Moen, 2003; Ackerly, 2003; Sargent & Ackerly, 2008; Hillebrand & Matthiessen, 2009). As a result of this concerted effort, a considerable number of papers addressing this broad question have been published (see online supporting information: Appendix S1 and S2). However, conclusions have tended to be inconsistent. Moreover, the diversity of approaches used means that a common practical framework is lacking.

This review has three main objectives. First, we present an overview of the definitions, methodological issues, and theoretical background connected with the study of assembly rules in plant communities. Second, we review and analyze the relevant published literature to classify the different approaches used and summarize the current state of

knowledge. Finally, drawing together the first two sections, we discuss the theoretical significance of the work carried out to date. This critical synthesis allows us to outline future research directions and provide guidelines that, we believe, could contribute to a better understanding of plant community assemblages.

(2) Defining assembly rules

Attempts to quantify community structure as deviation from randomness began long before the introduction of the term 'assembly rules'. In the 1950s and 1960s "analysis of interspecific associations" was introduced in quantitative plant ecology (Greig-Smith, 1957; Kershaw, 1964). For example, Greig-Smith (1983) tried to deduce the degree of biotic organization in the tropical forests of Trinidad from the proportion of significant "interspecific associations", calculated on the basis of species co-occurrence. Similarly, Pielou (1972) quantified the relative importance of positive and negative associations in a very similar fashion to later tests of assembly rules.

The term 'assembly rule' was introduced to ecology by Diamond (1975) who identified several instances of 'forbidden species combinations' among fruit-eating birds living in the New Guinea archipelago. He concluded that interspecific interactions, and particularly competition, lead to nonrandom co-occurrence patterns. However, his approach was challenged by Connor & Simberloff (1979) who claimed that many of the patterns attributed to interspecific competition could also be explained by random colonization of communities. A series of discussion papers with methodological emphases followed from this work (*cf.* Gotelli & Graves, 1996; Fox, 1999 for a review).

While the original assembly rules referred only to biotic restrictions on observed patterns of species assemblage - a view also adopted by Wilson (1999*b*) - other authors considerably broadened the scope of the concept. Keddy (1992) defined assembly rules as any ecological process selecting for or against species from the regional species pool, and thus determining local community composition. These selective processes have been equated conceptually to hierarchical "filters" that act at increasingly finer scales to impose rules on the assemblage of communities (Fig. 1). Meanwhile, in his "unified neutral theory of biodiversity" Hubbell (2001) distinguished between niche-assembly and dispersal-assembly rules, arguing that biodiversity patterns of trees in forests are mostly explained by dispersal processes alone.

For the sake of generality, we use the term 'assembly rules' herein to refer to any constraint on species coexistence. We share the view of Wilson (1999*a*) that, although the term "rules" may indicate otherwise, assembly rules refer to "restrictions on the observed patterns" in the data, and not to the processes that produce these patterns. We further distinguish between phylogeographic and ecological assembly rules. The term phylogeographic assembly rule denotes restrictions on observed assemblages due to historical patterns of speciation and migration. The term ecological assembly rule denotes restrictions on community structure and composition due to any of the following ecological filters: (1) dispersal, (2) the abiotic environment, and (3) biotic interactions (Fig. 1). To reflect these different filters we use the terms dispersal assembly rules, abiotic assembly rules, and biotic assembly rules, respectively. While all of these different assembly rules can be influenced by the phylogenetic relatedness of species in communities, we do not directly consider this point in the present review (see Emerson & Gillespie, 2008; Cavender-Bares *et al.*, 2009 for recent overviews). Moreover, studies that explicitly assess dispersal assembly with null models are scarce. Hence, the major focus of the present review is on biotic assembly rules. It should be noted, however, that many studies designed to record biotic assembly rules have in fact revealed the importance of abiotic rules.

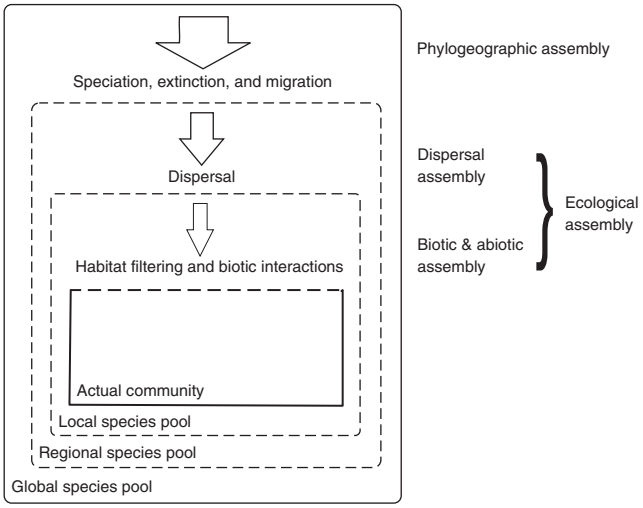


Fig. 1. The different processes that might produce assembly rules and the relative scales at which they are most influential (adapted from Zobel, 1997). At any point in time there is a global species pool that defines a regional species pool through the speciation, extinction and migration of species (phylogeographic assembly). At a given local site the species pool constitutes species from the regional species pool that are able to disperse there (dispersal assembly). At the local site, habitat filtering and biotic interactions define the actual assemblage of plant species (ecological assembly).

(3) The null model approach

Since the early work by Connor & Simberloff (1979), the study of assembly rules has been associated with the null model approach, which aims to identify the nonrandom component in community composition by comparing certain parameters of an observed data set with the same parameters in multiple randomized data sets. Random samples are generated using an assumed model - i.e. a mathematical formula or algorithm to redistribute the data - that is thought to represent the "null model". The term "null model" refers to the fact that the model is considered to represent the null

hypothesis that the observed pattern is a product of chance alone.

When addressing nonrandom patterns on the basis of species co-occurrence matrices, nine simple null models can be constructed using only the information contained in the row and column totals (Gotelli, 2000). These models differ in whether rows and columns are treated as fixed sums, equiprobable, or proportional, and differ in the probability of Type I and Type II errors occurring (see also discussion by Manly, 1995; Ulrich, 2004; Navarro-Alberto & Manly, 2009; Ulrich, Almeida & Gotelli, 2009). The design of appropriate null models becomes more complicated if information on species abundances and species traits is also incorporated (e.g. de Bello *et al.*, 2009). The construction of the randomization scheme under the null model is not merely a technical question; its definition determines which ecological mechanisms are allowed and which are excluded under the null model. Only those that are excluded can generate the deviation from the null model, but the criteria for inclusion can be subtle. For example, if the total biomass in a plot is fixed, the null model does not completely exclude competition, because the upper limit for biomass in a plot, reflected in the observed plot totals, is undoubtedly determined by competition. Hence, the choice of which margin in a community matrix is kept fixed when randomizing reflects different null hypotheses about the functioning of assembly rules.

(4) Spatial, ecological, and statistical dispersion

Assembly rules are assessed using data on spatial and/or ecological (i.e. trait) dispersion, i.e. determining whether species associate/dissociate spatially and whether ecological similarity between species is higher or lower than expected by chance. However, this is tested in a statistical framework that relies on measuring the dispersion of the data itself. This has generated a great deal of confusing terminology. In particular, the terms 'overdispersion' and 'underdispersion' have been used to describe spatial, statistical and ecological (i.e. species traits) dispersion patterns (see also Holdaway & Sparrow, 2007; Pausas & Verdu, 2010). However, there is no logical relationship between these terms when applied in these different contexts. Spatial overdispersion does not necessarily lead to overdispersion of trait data. Even more confusingly, trait overdispersion is reflected by statistical underdispersion in the data, and *vice versa*. We therefore follow the suggestion of Greig-Smith (1983) to avoid these terms completely, except where they are used in the sense of statistical dispersion. For spatial dispersion we use 'aggregation' and 'segregation', and for ecological dispersion we use trait 'convergence' and 'divergence' (see Section II.2).

(5) Spatial scale and variability

That scale is a crucial point to be considered by ecologists has long been recognized (Wiens, 1989; Levin, 1992). When investigating assembly rules in plant communities it is important to consider several issues related to the spatial scales at which assembly rules are expected to occur and at which they are measured. In general, it is important to define spatial scale precisely in terms of its two components: grain and extent. While grain defines "the size of the individual units of observation" (Wiens, 1989, p. 387), extent describes the overall area in which the units of observation are located. In plant community ecology these two terms generally equate to the sample plot size (or quadrat size) and to the area of the study site. The results of tests performed to detect assembly rules can depend on the size of the grain used, as well as on the spatial extent and ecological variability within this area (e.g. Swenson & Enquist, 2009). The intra- and interspecific interactions that potentially drive biotic assembly rules mostly act at small spatial scales (Stoll & Weiner, 2000), ranging from the immediate neighbours within a few centimeters or even millimeters up to a less-defined scale, maybe spanning one square meter or less for herbaceous plants. However, when considering assemblages of trees in forest communities, this upper limit is certainly higher (Wilson, 2000). Nevertheless, biotic assembly rules are expected to be mainly apparent at relatively small spatial scales (Bycroft *et al.*, 1993). Furthermore, if larger spatial scales containing environmental or fragmentation gradients are considered, any patterns arising due to biotic assembly rules may be confounded with patterns generated by abiotic (Gotelli & McCabe, 2002; Watkins & Wilson, 2003) and dispersal assembly rules.

When constructing a null model, it is important to consider whether spatial arrangement within the community matrix should be reflected by constraints on randomization. For certain community structure indices, patterns that are thought to be the result of biotic assembly rules could also be produced by environmental variability, i.e. abiotic assembly. This applies both at small scales, where it is referred to as microhabitat variability (Wilson *et al.*, 1987; Zobel & Zobel, 1988), and at larger scales, where habitat filtering may be acting (Cornwell, Schwilk & Ackerly, 2006). Attempts have been made to control for such variability by using environmentally or spatially constrained null models. In these null models, species are not randomized across the whole community matrix but only within neighbouring plots (i.e. the patch model of Watkins & Wilson, 1992; Peres-Neto, Olden & Jackson, 2001).

II. NICHE-COEXISTENCE THEORY AND DIFFERENT APPROACHES TO DETECTING ECOLOGICAL ASSEMBLY RULES

The study of assembly rules allows the mechanisms underlying species coexistence to be understood by linking observed community patterns to ecological interactions such as competition.

An interactive community is one in which strong biotic interactions take place among species at the same trophic level within a local habitat. Traditionally, interactive communities held a central place in the development of

community theory (Cornell & Lawton, 1992). According to this theory, niche space is filled, resulting in strong interactions between species on the same trophic level, either directly or indirectly *via* shared resources. Coexistence in the presence of these interactions is the major problem encountered by members of these communities. According to classical theory, species coexistence is possible due to differentiation in resource use (MacArthur & Wilson, 1967; MacArthur, 1972; Schoener, 1986). More recent work has proposed a number of stabilizing and equalizing mechanisms that might prevent competitive exclusion from occurring and thus make long-term coexistence of species possible (reviewed by Chesson, 2000). These studies have mostly addressed the coexistence of species which are already present in a community and not the role of dispersal or the availability of species in the local and regional species pools.

Different aspects of assembly rules have been addressed in relation to plant communities. Tests have traditionally focused on the taxonomic identity and diversity of species in communities, but an alternative approach is to consider the functional characteristics of species (functional traits) inhabiting these communities. Functional characteristics describe the adaptation of species to the environment, and taken together they reveal ecological differentiation between species. As such, they represent one of the most relevant components of biodiversity that can be considered in the context of community assembly rules. The following subsections survey the major approaches used to assess ecological assembly rules, as defined above (summarized in Table 1).

(1) Species-based approaches

(*a*) *Co-occurrences*

The presence–absence matrix is the fundamental unit of analysis in community ecology. In classical community theory, species co-occurrence is typically predicted to be nonrandom and less frequent than would be expected if species colonized sites entirely independently of one another (Gotelli & Graves, 1996). This prediction can be tested using a co-occurrence index, which is a single metric that summarizes the co-occurrence pattern in a presence–absence matrix (Gotelli, 2000). Frequently used examples of such indices are the C-score (Stone & Roberts, 1990) and the CHECKER index (Diamond, 1975). All co-occurrence indices detect spatial segregation or spatial aggregation of species. However, there are two problems concerning the presentation and interpretation of co-occurrence indices. Firstly, while co-occurrence studies often claim to demonstrate less co-occurrence than was expected under the null model, what is actually being calculated is the mean average pairwise co-occurrence of species or occurrence of the checkerboard pattern. Thus, the commonly used co-occurrence indices are interpreted as measuring co-occurrence *per se*, when actually they do not. For instance, consider a non-randomly constructed presence-absence matrix of eight sites containing eight species with a perfect checkerboard distribution. If co-occurrence is assessed by counting aggregated and segregated species pairs, 12 aggregated and 16 segregated species pairs would be recorded. Assessing this pattern with different co-occurrence indices would however produce different numbers (e.g. C-score = 9.14). Hence, it seems ambiguous still to interpret this in terms of co-occurrence.

Secondly, while segregation is most often interpreted in terms of competition, it can also arise from habitat variability. Similarly, aggregation might arise both from habitat variability and from facilitation, i.e. beneficial rather than competitive interspecific interactions (Callaway & Walker, 1997; Kikvidze *et al.*, 2005; Dullinger *et al.*, 2007). Thus, the results of co-occurrence indices tend to be ambiguous if studies do not control for environmental heterogeneity. Where such control is not possible, focusing on certain species pairs of interest might be more meaningful.

(*b*) *Niche limitation*

An alternative approach is to address the number of coexisting species, since classical community theory predicts that the number of species in a community is limited by the number of niches, as determined by the local environment - niche limitation (Wilson, Gitay & Agnew, 1987). Evidence for niche limitation can be tested by comparing the observed variation in species richness or diversity to that in randomly assembled communities (Palmer, 1987; Wilson *et al.*, 1987; Zobel & Zobel, 1988). Lower than expected variance in these indices is thought to indicate niche limitation—i.e. strong biotic interactions limiting the co-occurrence of more species in a given habitat. Higher than expected variance is interpreted as an indication of environmental heterogeneity, which in turn, implies the existence of abiotic assembly rules.

(2) Trait-based approaches

(*a*) *Guild proportionality*

Incorporation of abundance and trait data into the community composition matrix makes it possible to address the mechanisms of species coexistence with greater precision. For instance, it becomes possible to assess the degree of constancy in the number and proportion of species not simply throughout the whole plant community, but also within guilds, where the level of competition is expected to be greatest (Wilson, 1989). The term guild has been used for a group of species that are, or are assumed to be, similar in an ecologically relevant way (Wilson, 1999*b*). In a wider context than that of plant communities, one might for instance similarly consider trophic levels or broad taxonomic groups.

(*b*) *Limiting similarity*

Besides addressing the guilds in communities, resolution can be further increased by studying the dissimilarity between

Table 1. Overview of the major approaches used to study ecological assembly rules in plant communities. All approaches address the same general 'working hypothesis': that the amount of resources (the number and size of niches) is limited and that biotic interactions (notably interspecific resource competition) are the main force structuring plant communities. The approaches differ with respect to the community parameters addressed

Approach	Short description of the (competition-induced) pattern addressed	Community data	Trait data	Typical test
Co-occurrence	Competitive exclusion reduces the number of possible species co-occurrences	Presence-absence matrices	Not needed	Test whether there are fewer species co-occurrences than expected by chance
Niche limitation	The restricted number of niches and competitive exclusion limits the number of species (or amount of biomass) that can coexist in the same patch	Presence-absence or abundance matrices (cover, biomass)	Not needed	Test whether the ratio of observed to expected variance in species diversity is lower or higher than expected under the null model
Guild proportionality	Because competitive exclusion occurs mainly within functional guilds, the number of species within different guilds (i.e. occupying similar niches), or the total abundance of a particular guild in a community, are constant among patches.	Presence-absence or abundance matrices (cover, biomass)	Groups of species, usually categorical trait data (life form etc.)	Test whether the ratio of observed to expected variance in guild proportions is lower or higher than expected under the null model
Limiting similarity	Traits of coexisting species are more dissimilar due to the avoidance of competitive exclusion through niche differentiation. However, at large spatial scales and in heterogeneous environments, one may detect more similarity among species inhabiting similar environmental conditions (habitat filtering).	Species lists, presence-absence or abundance matrices (cover, biomass)	Usually continuous variables reflecting functionally important traits	Test whether trait values in a community are more dissimilar or more similar than expected under the null model

species in the community in terms of functional traits. The term limiting similarity stems from the competitive exclusion principle of Gause and was introduced by MacArthur & Levins (1967) to describe the maximum similarity in resource-use patterns that is consistent with the coexistence of two or more competing species. Within the context of assembly rules, the concept of limiting similarity suggests that species assembly within communities is arranged according to resource-use characteristics and that species can more readily coexist if they differ in their traits, such that competition between them is reduced. It is thus reasonable to test whether variation in the traits of coexisting species is different from that expected by chance, with trait divergence interpreted as evidence of limiting similarity (Watkins & Wilson, 2003). The opposite pattern, trait convergence, is interpreted as habitat filtering, which reflects the assemblage of species with shared ecological tolerances (Cornwell *et al.*, 2006; Ackerly & Cornwell, 2007), and thus represents an abiotic assembly rule. However, these patterns depend on the traits chosen for analysis (e.g. whether they are under environmental selection or not) and the scale at which they are measured, since detection of trait convergence or divergence in analyses may be strongly scale dependent. Finally, within relatively homogenous environmental conditions, trait convergence could also be caused by the exclusion of species bearing traits associated with low competitive abilities (Mayfield & Levine, 2010).

The study of limiting similarity is in fact analogous to the comparison of observed *versus* expected functional diversity, where high functional diversity is defined as high dissimilarity between species in the values of one or several traits (Petchey & Gaston, 2002; Leps *et al.*, 2006). Another aspect is the assessment of the relevance of different community assembly rules (Cornwell *et al.*, 2006;

Petchey *et al.*, 2007; Mason *et al.*, 2008*a*,*b*; De Bello *et al.*, 2009). This usually follows an approach already applied in species and phylogenetic diversity studies (Gotelli & McCabe, 2002; Kraft *et al.*, 2007) where the functional diversity value of an observed community is compared with those derived from multiple null communities, generally resulting from randomization of species composition across samples.

III. GLOBAL SURVEY AND SYNTHESIS

(1) Data collection and meta-analyses

Problems related to the detection of non random patterns of community assembly were assessed by reviewing relevant literature, with particular attention paid to the scale of the sampling design and the conceptual approach used (see Table 1).

In total we surveyed 59 papers published between 1987 and 2010, which reported 9658 trials comparing observed patterns with null models representing random patterns (Table 2, Appendix S1 and S2). Papers were chosen by searching in the *ISI - Web of Science* with the terms (“community assembly” OR “assembly rule*”) AND (“null” OR “random”). From the results we narrowed our selection to papers dealing with plant communities that addressed ecological assembly rules using null model testing. Some papers that were not found using these search terms but that clearly dealt with assembly rules in plant communities were added from personal literature databases. While we cannot guarantee that we detected all suitable papers, the reference list is, to our knowledge, the most comprehensive assessment to date of assembly rules in plant communities, and we expect it to serve as a point of reference for future work.

From the literature we recorded two types of response variables for use in our analyses. First, we assessed the number of tests in each paper and the proportion of these tests that deviated from the null model. Second, we recorded data relating to the actual indices of community structure. These consisted of standardized effect size (SES) values for the C-score and CHECKER indices from co-occurrence studies and relative variance index (RV) values from studies of niche limitation, guild proportionality and limiting similarity. The collection and analysis of the two types of data is described in the following two subsections.

(*a*) *Proportional data*

For each paper we counted the number of null model tests that were conducted, the number of tests where a deviation from the null model was observed and the direction of the deviation from the null model (i.e. towards biotic or abiotic assembly). Where possible, the grain size and extent at which null models were tested was also recorded. Since few studies report the actual extent of the study area, an approximate extent was calculated by multiplying the grain size with the number of replicate plots, resulting in the minimum area covered by a study. Some authors also distinguish interval (sometimes called focus) as a further component of spatial scale (Fortin & Dale, 2005). Focus represents the spatial distance between grains (plots, quadrats) but is not given by most of the studies and cannot be estimated. Therefore, we do not consider it in our analyses.

Additionally, we recorded the absolute latitude where the study was conducted, whether community structure indices were weighted by abundance measures (presence-absence weighted, cover value weighted, biomass weighted, frequency weighted), what plant types were used in the study (lower plants, higher plants, lower and higher plants combined, woody plants), and a broad habitat category (grassland, forest, wetland, old field, submersed). For each study, data were aggregated according to the grain size, the abundance measure and use of the patch model, i.e. counts of all tests and those tests deviating from the null model were summed for each combination of categories prior to further analysis. Some papers were excluded from this assessment because they used approaches that we could not clearly assign to one of the four approaches represented in Table 1 or deviations from the null model were not clearly interpretable in terms of biotic *versus* abiotic assembly rules.

It should be noted that the 9658 trials were not evenly distributed among the studies. Reflecting the design of the study, some papers only reported two or three trials, while others reported several hundreds. Furthermore, different scales, traits, null models and approaches were sometimes used within the same study, leading to a complicated data structure and a fairly unbalanced overall dataset. However, this reflects the complex nature of assembly rule null model studies and does not prohibit analysis of existing data to look for general patterns.

Table 2. The number (N) of single null model tests in the literature and the number of significant deviations from the null model among these tests. Deviations are further separated into deviations in the direction of biotic or abiotic assembly. The numbers and percentages are grouped according to their conceptual approach (see Table 1)

	N of papers	N of tests	N of significant deviations	Direction of biotic assembly	Direction of abiotic assembly	% significant	% biotic	% abiotic
Co-occurrence	11	3681	1501	436	1065	41%	29%	71%
Niche limitation	20	976	495	384	111	51%	78%	22%
Guild Proportionality	17	3035	482	227	255	16%	47%	53%
Limiting similarity	21	1966	358	229	129	18%	64%	36%
Total		9658	2836	1276	1560	29%	45%	55%

The effects of grain size, extent, latitude, plant types, abundance measure and habitat on the proportion of tests deviating from the null model were analyzed using generalized linear models with a quasi-binomial error distribution to account for overdispersion in the model. The proportion of deviations from the null model per study and scale was the response variable. Because the data were unbalanced and variables intercorrelated, we used univariate models for each of the variables rather than fitting a single model that included all explanatory variables. For each explanatory variable we fitted two models: one for the proportion of tests deviating in the direction of biotic assembly rules; the second for deviation in the direction of abiotic assembly rules. Grain size and extent were log-transformed for all analyses. Several studies used different spatial grain sizes in their sampling scheme. For these studies we fitted a generalized linear mixed model [GLMM; calculated using the *lme4* package (Bates, Maechler & Bolker, 2011) for R (R Development Core Team, 2010)] with only grain size as the explanatory fixed variable and publication as the random variable, to account for the group structure within publications.

(*b*) *SES and RV values*

For indices of species co-occurrence (i.e. C-score and CHECKER values), an "effect size" can be calculated by standardizing the difference between "control" and "treatment" groups (Gurevitch *et al.*, 1992). We followed the approach of Gotelli & McCabe (2002) and used the standardized effect size (SES), which measures the number of standard deviations that the observed index is above or below the mean index of the simulated communities. SES was recorded when provided by the authors or calculated from the C-score or CHECKER values (see Gotelli & McCabe, 2002) as:

$$\mathrm{SES} = (I_{\mathrm{obs}} - I_{\mathrm{sim}})/\sigma_{\mathrm{sim}}, \quad (1)$$

where I_{obs} is the co-occurrence index of the observed community matrix, and I_{sim} and σ_{sim} are the mean and the standard deviation of the indices simulated under the null model, respectively. The SES of these co-occurrence indices has a mean of 0 under the null hypothesis, with values above 0 indicating species segregation (biotic assembly) and values below 0 indicating species aggregation.

Many papers studying niche limitation or guild proportionality have used the relative variance index (RV) introduced by Wilson (1989). RV is calculated as

$$\mathrm{RV} = V_{\mathrm{obs}}/V_{\mathrm{sim}}, \quad (2)$$

where V_{obs} is the variance in species richness (niche limitation) or guild proportionality of the observed community matrix, and V_{sim} is the variance in species richness or guild proportionality of the simulated community matrices. RV values can vary from 0 to infinity, with values below 1 indicating biotic assembly rules, and values above 1 indicating abiotic assembly.

For both SES and RV values, we analyzed variation in relation to grain size, extent, latitude, studied plant types, and habitat. In addition, we analyzed variation in relation to null model type (i.e. if both column and row margins were fixed for randomization, or only column margins) for SES data and in relation to use of a patch model for RV data. Because SES and RV values were not normally distributed we used Kendall's correlation tests (continuous variables) and Kruskal-Wallis rank sum tests (categorical variables). The overall distributions of SES and RV values were analyzed with Wilcoxon rank tests to check if the values were distributed symmetrically around the hypothetical centers of 0 (SES) or 1 (RV), since none of the distributions were normally distributed according to Shapiro-Wilk tests. All statistical tests were carried out using the R statistical software, version 2.11.1 (R Development Core Team, 2010).

It should be noted that we did not challenge the validity of the statistical and methodological approaches, nor of the underlying theoretical assumptions, related to the null models used in the analyzed papers. Therefore, we assessed the results as they were presented by authors. We further provide explanations for the potential biases that could lead to Type I and/or Type II errors within each of the four mentioned approaches. Despite such possible bias, the standardized approach we followed for our meta-analyses enables us to be more confident about our conclusions than if they were based solely on a narrative review.

(2) Species-based approaches

(*a*) *Co-occurrences*

There were 11 papers reporting 3681 trials to test the randomness of plant species co-occurrence patterns in communities. Significant deviations were found in 41% of all trials; of these 29% indicated patterns that could be attributed to biotic assembly, and 71% indicated species aggregation that could reflect either facilitation or abiotic assembly (Table 2). Accordingly, the SES for C-score values were significantly below zero (Fig. 2A; mean SES = −1.26, median SES = −0.72, $V = 1368025$, $N = 2977$, $P < 0.001$), indicating deviation from randomness in the direction of species aggregation. However, this pattern was dependent upon the SES values from a high number of alpine vegetation plots ($N = 1808$) from one study (Dullinger *et al.*, 2007; see Fig. 2). The SES for CHECKER values were not significantly different from zero (Fig. 2B; mean SES = −0.18, median SES = −0.14, $V = 42961$, $N = 463$, $P = 0.0024$). These results show that non-random co-occurrence (in terms of checkerboard units) of plant species in the direction of competition-driven biotic assembly was infrequently demonstrated by the studies included in our analysis. Most SES values were either not different from random or were indicative of either facilitation or abiotic assembly.

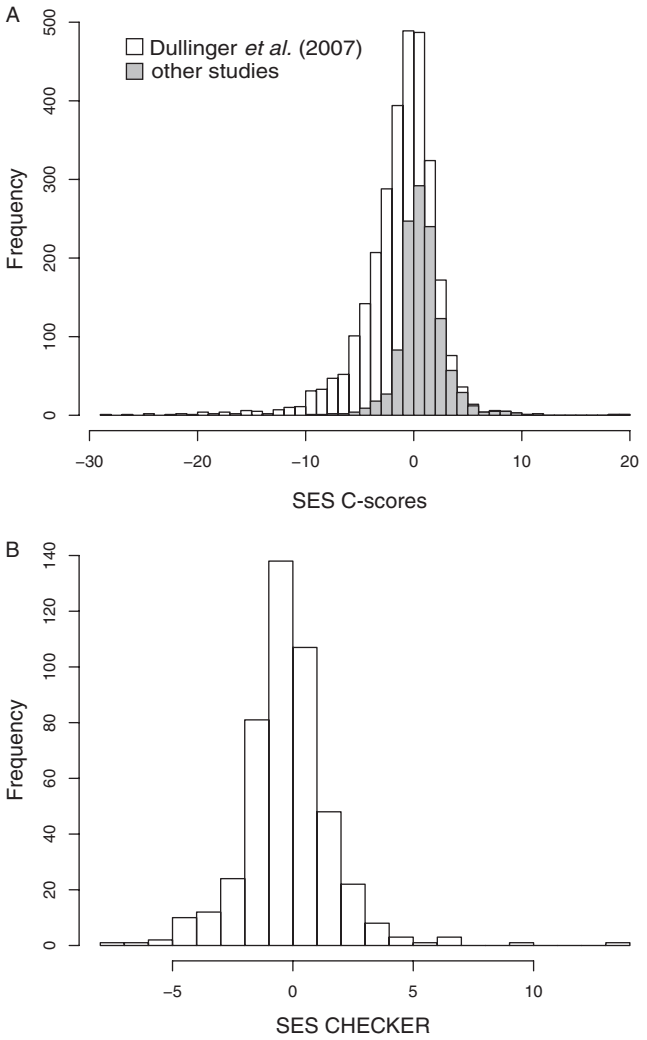


Fig. 2. Distributions of standardized effect size (SES) values for (A) C-scores and (B) CHECKER values.

By contrast, Gotelli & McCabe (2002) found "less co-occurrence" in real matrices than expected by chance, matching the predictions of Diamond's (1975) model. They analyzed 13 published plant co-occurrence matrices that had not been previously analyzed in the context of assembly rules, and their data set does not overlap with that used in our meta-analyses. However, as those authors mentioned, the main caveat of this approach is that results may also be influenced by a 'habitat checkerboard' effect—i.e. segregation of species because of their affinities for non-overlapping habitats. The same might be true for some of the studies included in our analysis, and thus no clear conclusions can be drawn at this point.

(*b*) *Niche limitation*

Twenty reviewed papers addressed the concept of niche limitation (Table 2). About half of the tests reported on niche limitation showed deviation from the null model, with 78% of these significant deviations attributable to niche limitation, and 22% indicative of greater variation of richness or diversity than expected by the null model. RV values from six studies from which we could extract the necessary data showed no significant difference from 1 (mean RV = 1.06, median RV = 0.98, $V = 4507$, $N = 134$, $P = 0.97$). However, the RV values were not distributed normally and a high proportion of the RV values were in fact below 1 (Fig. 3C), indicating less variation in richness or diversity than expected by the null model, and thus niche limitation. Also, the relatively high proportion (about two thirds; Table 2) of studies reporting niche limitation supports the view that the number of niches in plant communities is limited.

(3) Trait-based approaches

(*a*) *Guild proportionality*

Seventeen papers reported the results of 3035 null-model-based tests of guild proportionality in plant communities. The guilds considered were overwhelmingly either broadly defined growth forms or life-history types. Significant deviations from null models were recorded in less than a fifth (16%) of cases, with 47% and 53% of the significantly deviant tests indicating biotic and abiotic assembly, respectively. RV values were significantly different from 1 (Fig. 3B; $V = 4615.5$, $N = 186$, $P < 0.05$); the mean and median of the distribution, however, were close to 1 (i.e. 1.08 and 0.98, respectively), indicating no consistent deviation from randomness.

According to these results, guild proportionality seems not to be a commonly detected phenomenon in plant community assembly. This result may reflect the true nature of plant communities, but may equally reflect two methodological pitfalls. First, only three of the reviewed studies weighted guild proportionalities by the abundance of the species (Wilson & Gitay, 1995; Wilson *et al.*, 2000; Bossuyt, Honnay & Hermy, 2005); all others assessed guild proportionality based on the number of species in the guilds. Second, the predominantly morphological guilds addressed might not represent the primary 'functional units' of plant communities. Future studies should consider guilds not studied up to now. For instance, so far plants' ability to interact with different types of heterotrophic organisms has not been taken into account when addressing guild proportionality. One function to address in this context could be mycorrhizal status, i.e. whether plant species have the possibility to exploit resources more thoroughly through symbiotic interactions with mycorrhizal fungi.

(*b*) *Limiting similarity*

Altogether 21 papers reported 1966 tests comparing the dispersion of 63 plant traits against null models. Most traits were qualitative or quantitative traits characterizing the morphology and structure of above-ground vegetative organs or seeds (Table 3). Significant deviations from random were found in about 18% of cases, of which 64% reported higher than expected variation of traits, and 36% lower than expected variation of traits. Thus, in the majority of cases, the variation of traits of coexisting species did not differ from random,

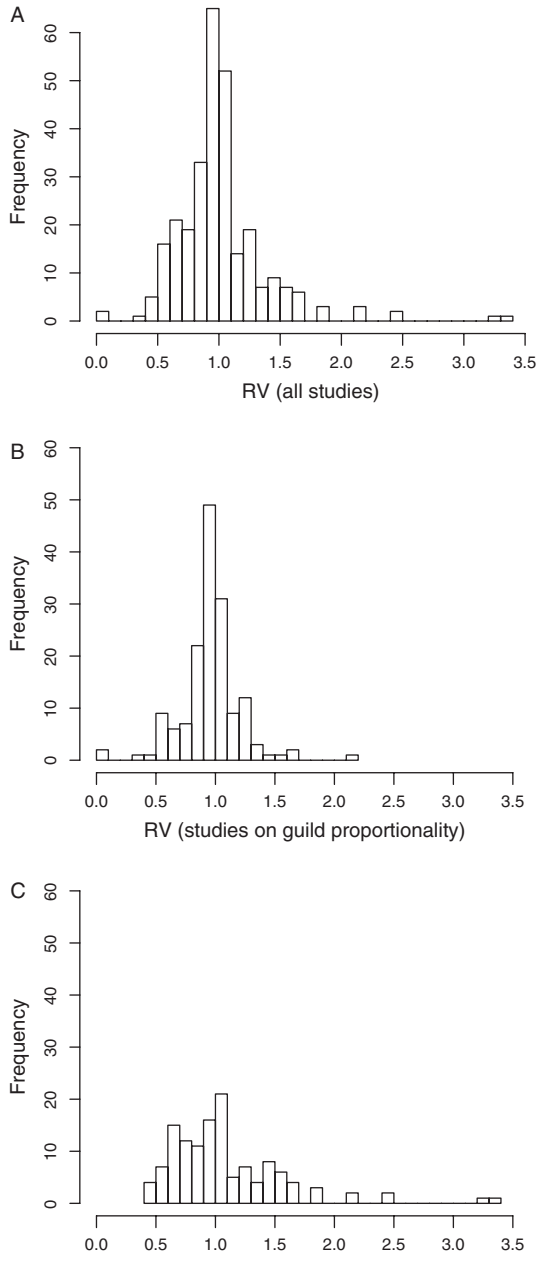


Fig. 3. Distributions of relative variance (RV) values (A) from all studies of guild proportionality and niche limitation, (B) only from studies of guild proportionality, and (C) only from studies of niche limitation.

but when significant deviations were found, trait divergence (reflecting 'limiting similarity') was more common than trait convergence (reflecting 'habitat filtering'). Although the same trait could show convergence, divergence or no significant pattern in different studies, there was a tendency for some of the traits used in limiting similarity studies—particularly leaf traits—to exhibit more frequent deviation from the null model than others (Table 3; in order of the overall proportion of deviant tests: specific leaf area, leaf area, leaf thickness, leaf succulence, and leaf nitrogen content). We also examined trait-related convergence and divergence separately for studies conducted in forest habitats and other habitats (mostly grasslands). For all of the traits where it was possible to make such a distinction, there was a clear tendency for more deviant tests in studies using data from forest habitats compared to studies of non-forest habitat types (Table 3).

The relatively high proportion (82%) of non-significant results might be partly due to methodological issues. When using a null model approach, biotic and abiotic filters may cause deviations in opposite directions from one another. It is thus entirely possible that the two effects compensate each other, such that the value of the test statistic supports the null hypothesis. In this way, apparently random assemblages could in fact result from a combination of convergence and divergence.

(4) Methodological considerations

(*a*) *The accuracy of the null model*

Key to successful application of the null model approach are selection of an appropriate model to test the specific biological question asked, and correct interpretation of a given departure of observed patterns from the expected patterns generated by the null model (Gotelli & Graves, 1996).

Analyzing pooled data from papers using the co-occurrence approach, we found that C-score and CHECKER SES values were significantly higher for null models that kept species frequencies over the plots and richness within plots constant, compared with those that kept only the species frequencies constant (Table 4, median fixed-fixed null model: 0.73, median fixed-equiprobable model: −1.60). This significant pattern remained when SES for C-scores and CHECKER values were analyzed separately, and also when values from one study (Dullinger *et al.*, 2007) that contributed about half of the data on SES values were removed. These results might suggest that the two null models detect different patterns in community matrices, but it has to be kept in mind that the interpretation of these patterns is often not straightforward (see Section II.1*a*). Regarding the analysis of RV values, we did not detect an influence of using a patch model for constructing randomized communities (Table 4).

Unfortunately, it was sometimes hard to determine the randomization procedure used in the different papers considered for the meta-analysis, because the authors did not provide a transparent description of their methods and underlying assumptions. The scientific significance of such publications thus remains under question, and consequently the value of our meta-analysis is in documenting this uncertainty and highlighting the need for a more formalized framework to assess the existence of assembly rules. Even when the null model algorithm is thoroughly described, validating the probability of Type I and II errors using simple model data with a known structure, as done by Zobel, Zobel & Peet (1993) and Gotelli (2000), remains the exception rather than the rule.

Table 3. Overview of traits analyzed in the study of limiting similarity and the number of tests conducted within these studies, the percentage of tests deviating from the null model, and the percentages of deviant tests indicating biotic or abiotic assembly. Only traits considered in three or more papers are presented. Altogether 17 papers analyzed limiting similarity

Trait used in limiting similarity analysis	Number of papers	Number of tests	% deviant tests	% biotic (of deviant tests)	% abiotic (of deviant tests)
Leaf area	9	146	36%	66%	34%
forest habitat	5	62	55%	50%	50%
other	4	84	23%	95%	5%
Multivariate[1]	8	102	23%	62%	37%
forest habitat	3	15	47%	0%	100%
other	5	60	17%	80%	20%
Plant height	8	135	12%	29%	70%
forest habitat	3	11	36%	75%	25%
other	5	124	10%	15%	85%
Specific leaf area	8	131	38%	36%	64%
forest habitat	4	26	62%	44%	56%
other	4	103	32%	33%	67%
Seed size	5	54	9%	40%	60%
forest habitat	4	12	42%	40%	60%
other	1	42	0%	0%	0%
Leaf shape	5	55	5%	100%	0.0%
Leaf thickness	5	121	26%	81%	18%
forest habitat	1	36	58%	71%	29%
other	4	85	13%	100%	0%
Leaf succulence	4	85	20%	76%	23%
forest habitat	2	41	39%	75%	25%
other	2	44	2%	100%	0%
Leaf nitrogen content	3	64	18%	41%	58%
forest habitat	2	22	55%	42%	58%
other	1	42	0%	0%	0%
Support fraction	4	84	16%	100.0%	0.0%
Photosynthetic characters (chlorophyll content etc.)	4	120	5%	100%	0.0%
Leaf area ratio	3	49	10%	80%	20%
Wood density	3	21	47%	40%	60%

[1]Multivariate means that a metric has been calculated on the basis of several traits, i.e. reflecting multidimensional niche space.

(*b*) *Spatial scale and variability*

It is clear that the outcome of any null model approach based on plant community data is dependent on the scale at which the data were obtained (Weiher & Keddy, 1995). When analyzing the effect of grain size on the proportion of tests deviating from the null model in the direction of biotic assembly we found a significant relationship across all studies (Figure 4; Table 5, slope estimate = −0.093) and when accounting for correlation between observations

Table 4. Results of correlation analyses and Kruskal-Wallis tests of the relationship between standardized effect size (SES) and relative variance index (RV) values, and variables recorded from the reviewed literature

		SES			RV		
	Test	Statistic	N	P value	Statistic	N	P value
Grain size	Kendall's correlation (tau)	−0.053	3440	<0.001	0.212	291	<0.001
Extent	Kendall's correlation (tau)	−0.121	3414	<0.001	0.188	291	<0.001
Latitude	Kendall's correlation (tau)	0.034	3440	0.004	−0.060	291	0.145
Habitat	Kruskal-Wallis test	176.8	3440	<0.001	4.3	291	0.112
Plant type studied	Kruskal-Wallis test	75.7	3440	0.000	13.9	291	<0.001
Null model	Kruskal-Wallis test	938.9	3440	<0.001	—	—	—
Patch model	Kruskal-Wallis test	—	—	—	0.68	291	0.408
Abundance weighting	Kruskal-Wallis test	—	—	—	0.004	291	0.948

within publications by means of a mixed-effect model (Chi squared = 4.052, d.f. = 1, $P = 0.044$, slope estimate = −0.42). Hence, for our proportional data, biotic assembly rules are more often detected at relatively smaller grain sizes. However, no relationship was found between grain size and the proportion of tests exhibiting deviation in the direction of abiotic assembly. No relationship with extent was found for the proportion of tests in the direction of either biotic or abiotic assembly rules (Table 5). Because interactions between trees are expected to occur at larger scales than those between herbaceous plants, we repeated the analyses separately for studies that analyzed data on woody species and studies using data on herbaceous plants. This qualitatively changed two aspects of the results relating to woody plants; within studies on woody plants there was (1) a negative relationship between grain size and the proportion of tests deviating in the direction of abiotic assembly (Table 5; slope estimate = −0.40) and (2) a positive relationship between extent and the proportion of tests deviating in the direction of biotic assembly (Table 5; slope estimate = 1.04). Furthermore, SES of CHECKER and C-score values were negatively correlated with both grain size and extent, while RV values were positively related to grain size and extent (Table 4). Some studies have specifically considered the effect of spatial grain size, reporting either the scale dependence of assembly rules (Reitalu *et al.*, 2008; Zhang *et al.*, 2009), or no relationship (Dullinger *et al.*, 2007). The results of our analyses emphasize the general importance of scale for assembly rules studies and support the theoretical consideration that processes leading to assembly rules act at different scales for trees and herbaceous plants. However, because factors other than scale have a major influence on the results of single studies, we can not draw any more general conclusions regarding recommendations at what specific grain or extent sizes assembly rule studies should be conducted.

(*c*) *Further factors affecting the outcome of assembly rule studies*

The GLMs analyzing the single effects of latitude, habitat and plant type studied showed no significant effects (Table 5). Abundance weighting had a significant effect on the proportions of tests deviating from the null model in the directions of both biotic and abiotic assembly. The latter effect was not replicated in the analysis of RV values (Table 4), but here only two weighting categories (presence-absence and biomass) could be included due to the lack of data for the other categories. There was a significant positive effect of latitude on SES values, suggesting that biotic assembly rules might be more important in temperate habitats. However, the relationship was weak, and no similar effect was found for RV values (Table 4). In contrast to the GLM results, the plant type studied showed a significant relationship with SES and RV values. Pairwise Bonferroni-corrected Wilcoxon signed-rank tests between groups revealed that SES values from studies of higher plants (excluding bryophytes) were higher than those from studies on lower plants (bryophytes and lichens). None of the remaining pairwise comparisons were significant. For RV values, the data could only be split into two groups: studies on higher plants and studies carried out simultaneously on higher and lower plants. The significant difference between these two groups might indicate the importance of including bryophytes when analyzing community structure in habitats mainly made up by herbaceous plants. Differences in SES values among habitat types were caused by relatively high values reported in studies on grasslands (median = 0.83) and one study in submerged vegetation (median = 4.70), while medians for the other habitats were either close to zero (forests: −0.16; outcrops: 0.03) or smaller (multiple

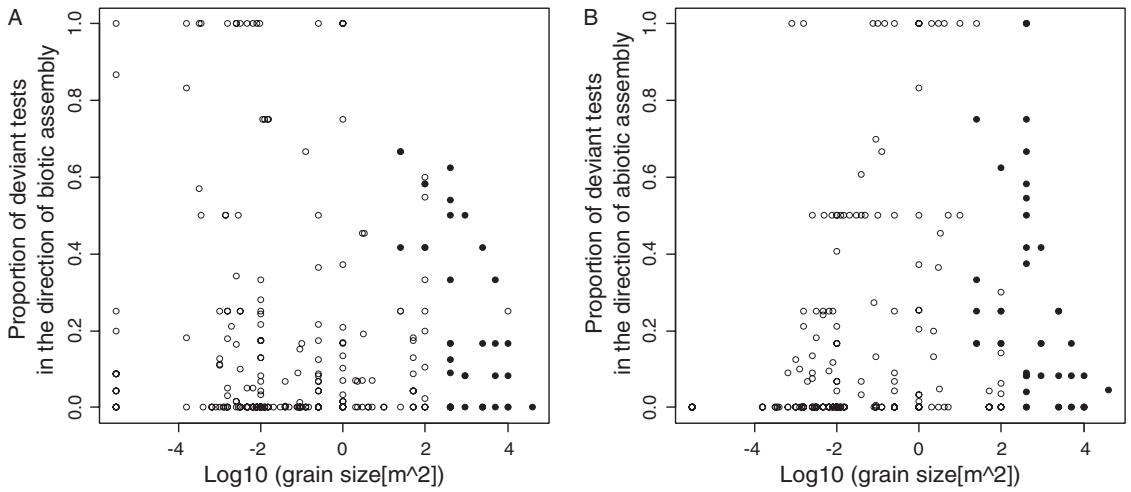


Fig. 4. Relationship between the grain size of a study and the emergence of assembly rule patterns. Each point represents the proportion of tests deviating from the null model in a certain study at a particular grain size. (A) The relationship for deviations in the direction of biotic assembly; (B) the relationship for deviations in the direction of abiotic assembly. Filled circles show studies that used data on woody species, open circles show other studies.

Table 5. Results of univariate generalized linear models (GLM) analyzing the relationship between the proportions of tests deviating from the null model in the direction of biotic and abiotic assembly, and variables recorded from the reviewed literature

	Biotic			Abiotic		
	Chi squared	d.f.	*P* value	Chi squared	d.f.	*P* value
Grain size	40.331	1	<0.001	0.214	1	0.644
Studies on herbaceous plants	44.473	1	<0.001	2.590	1	0.108
Studies on woody plants	58.254	1	<0.001	117.190	1	<0.001
Extent	0.660	1	0.417	0.325	1	0.568
Studies on herbaceous plants	0.359	1	0.549	0.053	1	0.818
Studies on woody plants	5.997	1	<0.001	0.536	1	0.464
Latitude	3.509	1	0.061	1.160	1	0.281
Habitat	4.337	8	0.826	7.102	8	0.526
Plant type studied	0.821	3	0.845	5.969	3	0.113
Abundance weighting	11.279	3	0.010	21.460	3	<0.001

habitats in one study: −0.77). The analysis of SES values was conducted excluding values from Dullinger *et al.* (2007). When including this study, medians of SES values for studies on higher plants and grasslands were markedly lower (−0.74 and −0.89, respectively).

Although these results suggest some general trends, it has to be kept in mind that inferring processes from SES values can be hindered by misinterpretation (see Section II.1*a*) and that we did not account for potentially nested data structures (e.g. traits within habitats).

In general, the influence of those factors considered in this section on the outcome of assembly rule studies indicates that generalizing the findings of a particular study is probably not justifiable. Apart from the spatial grain and extent, many other variables might determine the importance of assembly rules in a given plant community.

(5) Potential pitfalls in the study of assembly rules

The potential importance of niche differentiation received support from about a fifth of the case studies assessed in this review, while the remaining tests report either no pattern (about 70% of all tests) or deviation from random in the direction of abiotic assembly (16% of all tests). We see four potential reasons for the relatively low number of studies finding support for classical niche-based community theory.

Firstly, the role of niche differentiation in allowing plant coexistence may have been overestimated. Some authors have suggested that 'community ecology will have to rethink completely the classical niche-assembly paradigm' (Hubbell, 2001, p. 320). Indeed, the species composition and structure of plant communities is driven by multiple current and historical abiotic and biotic factors (Zobel, 1997; Ackerly, 2003; Lortie *et al.*, 2004; Agrawal *et al.*, 2007; Guisan & Rahbek, in press) and recent empirical evidence suggests that historic factors (Harrison & Grace, 2007; Zobel & Partel, 2008) and dispersal limitation (Zobel & Kalamees, 2005; Myers & Harms, 2009*b*) may be more important determinants of community composition than previously assumed. Given confounding factors (scale, type of ecosystem, disturbance, age of community and succession) and methodological issues, it remains difficult to draw general conclusions regarding the hierarchical importance of niche *versus* neutral *versus* historical processes. Nonetheless, this issue is fundamental to our understanding of how ecological communities are assembled.

Second, disentangling all possible causal relationships underlying actual community patterns on the basis of descriptive information alone is a great methodological challenge. It requires very clear working hypotheses and appropriate model systems, notably clear vegetation patterns, thorough information concerning species traits and appropriate scales of study. In particular, it is important for each individual study to address a complex of assembly rules and scales (Mason & Wilson, 2006). Addressing only a limited set of plant traits or a partial range of scale in a particular community for a limited time period might be insufficient to understand a potentially complex and fundamental phenomenon like species coexistence.

Thirdly, and related to the previous point, the diverse outcomes of studies searching for assembly rules is at least partly related to the diversity of methodological approaches, including the use of different study designs, data types (species occurrence or abundance), algorithms, statistical tests and dependent variables. Because methodological details are not always well described in publications (or described at all), it is currently impossible to evaluate what fraction of the variation in results is due to 'methodological noise' or even artifacts, and what represents true variation. A good way to make the approaches comparable would be to test them with simple model data and evaluate the rates of Type I and Type II errors associated with different approaches. Further meta-analyses that apply a common set of randomizations and statistics on a large number of community-matrices from published literature or vegetation databases might also help to draw more general conclusions.

Finally, while the study of assembly rules addresses patterns, inferring mechanisms from observed patterns has clear limits (Leps, 1990). Indeed, it is often hard or even impossible to disentangle the complex combination of processes that

shape plant communities. The quantification of their relative weight is challenging. Field and laboratory experiments are a promising alternative to observational studies. Up to now, there has been surprisingly little use of species removal and addition experiments, explicitly to study assembly rules (Fukami *et al.*, 2005; Turnbull *et al.*, 2005; Ejrnaes, Bruun & Graae, 2006; Mwangi *et al.*, 2007; Myers & Harms, 2009*a*; Roscher, Schmid & Schulze, 2009; McLaren & Turkington, 2010). The advantage of, e.g., removal experiments is that they should be able to disentangle the effects of habitat filtering, which should be unaffected by removal of species, and of biotic interactions in generating assembly rules. Assembly rule theory can provide strong predictions about the reaction of remaining species to the removal. For example, should limiting similarity be the principle shaping the plant community, then the removed species should be replaced by functionally similar species, or pre-existing functionally similar species should gain most from the removal. Hence, we strongly encourage further experimental approaches addressing assembly rules in plant communities.

(6) Guidelines for best practice

Despite the above-mentioned issues we are confident that null models can provide a strong tool for assembly rule research. Here, we present guidelines for best practice in order to encourage plant ecologists to embrace the application of null models, while bearing in mind the possible pitfalls:

(1) Though the randomization algorithm is one of the most crucial components of a null-model-based study, it is not possible to promote a particular algorithm (or even a small set of algorithms) for use in all studies. The choice of the null model should always be made by considering the processes, habitats, spatial scales, traits and organisms under investigation. As Gotelli (2000, p. 2619) pointed out when discussing null models applied to presence/absence data, ''ecologists need to move beyond the idea that there is a single 'one-size-fits-all' null model that is appropriate''. Thus, in any null model study, it is important to give a detailed description of data and the randomization procedure used; ideally documented with a flow chart. The appendix of Stubbs & Wilson (2004) provides a very good example. It should be clearly stated what mechanisms are excluded by the randomization procedure, and what deviations from the null model mean biologically.

(2) If feasible, more than one plot size (i.e. spatial grain size) should be considered. In the simplest case this can be done by lumping together adjacent plots of one size, but results at the different scales cannot then be considered independent. Ideally, results should be gathered from independent plots of different grain sizes.

(3) The spatial extent should be considered when randomizing community data. Appropriate extent should be defined by the hypothesized assembly rule. For instance, it makes little sense to test for biotic assembly by randomizing data over plots spanning a large study area, where environmental heterogeneity, and thus abiotic assembly, probably predominates. Assembly rule studies should always indicate the spatial extent and describe any environmental gradients in the study area; information that is often lacking in previous studies.

(4) In nature, plant individuals differ in size, and species differ in abundance. Meanwhile, dominant species are probably more important for ecological functioning than those that occur sparsely within the vegetation cover. Wherever possible, abundance data in the form of biomass, cover or individual counts should be preferred to presence/absence data.

(5) Data on functionally important plant traits are becoming increasingly available to the scientific community. However, such data may not be applicable for certain tasks, because they will almost certainly not reflect local conditions and may be aggregated over different localities. If possible, community data should be accompanied by trait data derived from individuals in the studied community. If databases are used for trait data, the suitability of the original data source and the quality of the data should be thoroughly checked.

(6) When using trait-based approaches, rather than screening many different traits, it is preferable to address primarily those traits that are functionally relevant to the assembly mechanisms addressed. While this might not always be a straightforward decision, cues to what functional processes are important in certain habitats or at certain scales can be found in the literature. For instance, if competition is assumed to be the important driving force of community assembly, traits related to competitive ability (Keddy, 2002; Wang *et al.*, 2010) should be primarily addressed. This step also encourages hypotheses to be devised prior to the study, rather then following an exploratory analysis.

IV. CONCLUSIONS

(1) The assembly rules approach investigates the mechanisms that structure biological communities. The approach goes to the heart of the old question of whether communities are formed by interacting species, and thus show non-random patterns (Clements, 1916), or represent random aggregations of species able to tolerate the same environmental conditions (Gleason, 1926). In particular, the search for biotic assembly rules addresses the most important stabilizing mechanism of species coexistence - niche differentiation. Reduced co-occurrence, niche limitation, guild proportionality and limiting similarity are all phenomena indicating, albeit from different angles, the presence of this mechanism.

(2) Here, we have reviewed the state of the literature as regards the existence and types of ecological assembly rules in plant communities. We found that non-random assemblage is only infrequently demonstrated in plant communities. However, the many possible biases and methodological shortcomings associated with published studies mean that our results do not constitute definitive evidence.

(3) This outcome can be considered disappointing in the sense that a long history of assembly rules study has failed to reach more definitive conclusions. However, by describing the current state of the art, we trust that this knowledge can be used to identify the most important steps to be taken next.

(4) Besides the more practical considerations presented in our guidelines for best practice, we need in the future new comparative studies that simultaneously address not only different categories of biotic assembly rules (co-occurrence, niche limitation, guild proportionality, limiting similarity), but also different types of assembly rules (abiotic, biotic, dispersal, even phylogeographic). This includes applying consistent sets of methods to different data sets, ideally combining different types of ecosystems, different plant traits and functional types, and studying patterns along a range of scales (i.e. from local to landscape to biogeographic). Furthermore, extending the approach from plant communities to incorporate different trophic levels, including plant-soil organism interactions (Bever *et al.*, 2010), should provide new insights.

V. ACKNOWLEDGEMENTS

We thank Stefan Dullinger, Fernando T. Maestre, Sara Maltez-Mouro, Triin Reitalu, Tom Rooney and Jian Zhang for providing raw data from their articles. The comments of J. Bastow Wilson and an anonymous reviewer on an earlier version of the manuscript greatly improved the quality of the paper. The following co-authors received funding from the funding bodies stated inside the parentheses: F.D.B. (IFB-ANR DIVERSITALP project, ANR 2008-2011, contract No. ANR 07 BDIV 014), J.L. (GACR 1374/206/09/1471), L.G., M.Z. & M.P. (European Union 7th framework project SCALES, FP7-226852; European Union 6th framework project ECOCHANGE, FP6-036866; European Regional Development Fund, Centre of Excellence FIBIR).

VI. REFERENCES

Ackerly, D. D. (2003). Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences* **164**, 165–184.

Ackerly, D. D. & Cornwell, W. K. (2007). A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. *Ecology Letters* **10**, 135–145.

Agrawal, A. A., Ackerly, D. D., Adler, F., Arnold, A. E., Caceres, C., Doak, D. F., Post, E., Hudson, P. J., Maron, J., Mooney, K. A., Power, M., Schemske, D., Stachowicz, J., Strauss, S., Turner, M.G. & Werner, E. (2007). Filling key gaps in population and community ecology. *Frontiers in Ecology and the Environment* **5**, 145–152.

Bates, D., Maechler, M. & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-38. http://CRAN.R-project.org/package=lme4

Belyea, L.R. & Lancaster, J. (1999). Assembly rules within a contingent ecology. *Oikos* **86**, 402–416.

Bever, J. D., Dickie, I. A., Facelli, E., Facelli, J., Klironomos, J., Moora, M., Rillig, M. C., Stock, W. D., Tibbett, M. & Zobel, M. (2010). Rooting theories of plant community ecology in microbial interactions. *Trends in Ecology and Evolution* **25**, 468–478.

Bossuyt, B., Honnay, O. & Hermy, M. (2005). Evidence for community assembly constraints during succession in dune slack plant communities. *Plant Ecology* **178**, 201–209.

Brooker, R. W., Callaway, R. M., Cavieres, L. A., Kikvidze, Z., Lortie, C. J., Michalet, R., Pugnaire, F. I., Valiente-Banuet, A. & Whitham, T. G. (2009). Don't diss integration: a comment on Ricklefs's disintegrating communities. *American Naturalist* **174**, 919–927.

Bruun, H. H. & Moen, J. (2003). Nested communities of alpine plants on isolated mountains: relative importance of colonization and extinction. *Journal of Biogeography* **30**, 297–303.

Bycroft, C. M., Nicolaou, N., Smith, B. & Wilson, J. B. (1993). Community structure (niche limitation and guild proportionality) in relation to the effect of spatial scale, in a Nothofagus forest sampled with a circular transect. *New Zealand Journal of Ecology* **17**, 95–101.

Callaway, R. M. & Walker, L. R. (1997). Competition and facilitation: a synthetic approach to interactions in plant communities. *Ecology* **78**, 1958–1965.

Cavender-Bares, J., Kozak, K. H., Fine, P.V.A. & Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters* **12**, 693–715.

Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics* **31**, 343–366.

Clements, F. E. (1916). *Plant succession: an analysis of the development of vegetation*. Carnegie Institution of Washington, Washington.

Connor, E. F. & Simberloff, D. (1979). The assembly of species communities: chance or competition? *Ecology* **60**, 1132–1140.

Cornell, H. V. & Lawton, J. H. (1992). Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. *Journal of Animal Ecology* **61**, 1–12.

Cornwell, W. K., Schwilk, D. W. & Ackerly, D. D. (2006). A trait-based test for habitat filtering: convex hull volume. *Ecology* **87**, 1465–1471.

De Bello, F., Thuiller, W., Leps, J., Choler, P., Clement, J. C., Macek, P., Sebastia, M. T. & Lavorel, S. (2009). Partitioning of functional diversity reveals the scale and extent of trait convergence and divergence. *Journal of Vegetation Science* **20**, 475–486.

Diamond, J. M. (1975). Assembly of species communities. In *Ecology and Evolution of Communities* (eds M. L. Cody and J. M. Diamond), pp. 342–444. Harvard University Press, Harvard.

Diaz, S., Cabido, M. & Casanoves, F. (1998). Plant functional traits and environmental filters at a regional scale. *Journal of Vegetation Science* **9**, 113–122.

Dullinger, S., Kleinbauer, I., Pauli, H., Gottfried, M., Brooker, R., Nagy, L., Theurillat, J. P., Holten, J. I., Abdaladze, O., Benito, J. L., Borel, J. L., Coldea, G., Ghosn, D., Kanka, R., Merzouki, A., Klettner, C., Moiseev, P., Molau, U., Reiter, K., Rossi, G., Stanisci, A., Tomaselli, M., Unterlugauer, P., Vittoz, P. & Grabherr, G. (2007). Weak and variable relationships between environmental severity and small-scale co-occurrence in alpine plant communities. *Journal of Ecology* **95**, 1284–1295.

Ejrnaes, R., Bruun, H. H. & Graae, B. J. (2006). Community assembly in experimental grassland: suitable environment or timely arrival? *Ecology* **87,** 1225–1233.

Emerson, B. C. & Gillespie, R. G. (2008). Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology & Evolution* **23**, 619–630.

Eriksson, O. (1993) The species-pool hypothesis and plant community diversity. *Oikos* **68**, 371–374.

Fargione, J., Brown, C. S. & Tilman, D. (2003). Community assembly and invasion: An experimental test of neutral versus niche processes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8916–8920.

Fortin, M. J. & Dale, M. (2005). *Spatial analysis - a guide for ecologists*. Cambridge University Press, Cambridge.

Fox, B. J. (1999). *The genesis and development of guild assembly rules. Ecological assembly rules* (eds E. Weiher and P. Keddy), pp. 23–57. Cambridge University Press, Cambridge.

Fukami, T., Bezemer, T. M., Mortimer, S. R. & Van der Putten, W. (2005). Species divergence and trait convergence in experimental plant community assembly. *Ecology Letters* **8**, 1283–1290.

Gleason, H. A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club* **53**, 7–26.

Gotelli, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology* **81**, 2606–2621.

Gotelli, N. J. & Graves, G. R. (1996). *Null models in ecology*. Smithonian Institution Press, Washington.
Gotelli, N. J. & McCabe, D. J. (2002). Species co-occurrence: A meta-analysis of J. M. Diamond's assembly rules model. *Ecology* **83**, 2091–2096.
Greig-Smith, P. (1952). Ecological observations on degraded and secondary forest in Trinidad, British West Indies: II. Structure of the communities. *Journal of Ecology* **40**, 283–330.
Greig-Smith, P. (1983). *Quantitative plant ecology*. Butterworths, London.
Grime, J. P. (2006). Trait convergence and trait divergence in herbaceous plant communities: Mechanisms and consequences. *Journal of Vegetation Science* **17**, 255–260.
Guisan, A. & Rahbeck, C. (in press). SESAM - A new framework for predicting spatio-temporal patterns of species assemblages: Integrating macroecological and species distribution models. *Journal of Biogeography*
Gurevitch, J., Morrow, L. L., Wallace, A. & Walsh, J. S. (1992). A meta-analysis of competition in field experiments. *American Naturalist* **140**, 539–572.
Harrison, S. & Grace, J. B. (2007). Biogeographic affinity helps explain productivity-richness relationships at regional and local scales. *American Naturalist* **170**, S5–S15.
Hillebrand, H. & Matthiessen, B. (2009). Biodiversity in a complex world: consolidation and progress in functional biodiversity research. *Ecology Letters* **12**, 1405–1419.
Holdaway, R. J. & Sparrow, A. D. (2006). Assembly rules operating along a primary riverbed-grassland successional sequence. *Journal of Ecology* **94**, 1092–1102.
Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton.
Keddy, P. A. (1992). Assembly and response rules: two goals for predictive community ecology. *Journal of Vegetation Science* **3**, 157–165.
Kembel, S. W. (2009). Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecology Letters* **12**, 949–960.
Kershaw, K. A. (1964). *Quantitative and dynamic ecology*. E. Arnold, London.
Kikvidze, Z., Pugnaire, F. I., Brooker, R. W., Choler, P., Lortie, C. J., Michalet, R. & Callaway, R. M. (2005). Linking patterns and processes in alpine plant communities: a global study. *Ecology* **86**, 1395–1400.
Kraft, N.J.B., Cornwell, W. K., Webb, C. O. & Ackerly, D. D. (2007). Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *American Naturalist* **170**, 271–283.
Leps, J. (1990). Can underlying mechanisms be deduced from observed patterns? *Spatial processes in plant communities* (eds F. Krahulec, A. D. Q. Agnew, S. Agnew and J. H. Willems), pp. 1–11. SPB Academic Publishing, The Hague.
Leps, J., De Bello, F., Lavorel, S. & Berman, S. (2006). Quantifying and interpreting functional diversity of natural communities: practical considerations matter. *Preslia* **78**, 481–501.
Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology. *Ecology* **73**, 1943–1967.
Lortie, C. J., Brooker, R. W., Choler, P., Kikvidze, Z., Michalet, R., Pugnaire, F. I. & Callaway, R. M. (2004). Rethinking plant community theory. *Oikos* **107**, 433–438.
MacArthur, R. H. (1972). *Geographical ecology. Patterns in the distribution of species*. Harper and Row, New York.
MacArthur, R. H. & Levins, R. (1967). The limiting similarity, convergence and divergence of coexisting species. *American Naturalist* **101**, 377–385.
MacArthur, R. H. & Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton University Press, Princeton.
Manly, B.F.J. (1995). A note on the analysis of species cooccurrences. *Ecology* **76**, 1109–1115.
Mason, N.W.H., Irz, P., Lanoiselee, C., Mouillot, D. & Argillier, C. (2008*a*). Evidence that niche specialization explains species-energy relationships in lake fish communities. *Journal of Animal Ecology* **77**, 285–296.
Mason, N.W.H., Lanoiselee, C., Mouillot, D., Wilson, J. B. & Argillier, C. (2008*b*). Does niche overlap control relative abundance in French lacustrine fish communities? A new method incorporating functional traits. *Journal of Animal Ecology* **77**, 661–669.
Mason, N.W.H. & Wilson, J. B. (2006). Mechanisms of species coexistence in a lawn community: mutual corroboration between two independent assembly rules. *Community Ecology* **7**, 109–116.
Mayfield, M. M. & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters* **13**, 1085–1093.
McLaren, J. R. & Turkington, R. (2010). Ecosystem properties determined by plant functional group identity. *Journal of Ecology* **98,** 459–469.
Mwangi, P. N., Schmitz, M., Scherber, C., Roscher, C., Schumacher, J., Scherer-Lorenzen, M., Weisser, W. W. & Schmid, B. (2007). Niche pre-emption increases with species richness in experimental plant communities. *Journal of Ecology* **95**, 65–78.
Myers, J. A. & Harms, K. E. (2009*a*). Local immigration, competition from dominant guilds, and the ecological assembly of high-diversity pine savannas. *Ecology* **90,** 2745–2754.
Myers, J. A. & Harms, K. E. (2009*b*). Seed arrival, ecological filters, and plant species richness: a meta-analysis. *Ecology Letters* **12**, 1250–1260.
Navarro-Alberto, J. A. & Manly, B.F.J. (2009). Null model analyses of presence-absence matrices need a definition of independence. *Population Ecology* **51**, 505–512.
Palmer, M. W. (1987). Variability in species richness within Minnesota oldfields: a use of the variance test. *Vegetatio* **70**, 61–64.
Pausas, J. G. & Verdu, M. (2010). The jungle of methods for evaluating phenotypic and phylogenetic structure of communities. *Bioscience* **60**, 614–625.
Peres-Neto, P. R., Olden, J. D. & Jackson, D. A. (2001). Environmentally constrained null models: site suitability as occupancy criterion. *Oikos* **93**, 110–120.
Petchey, O. L., Evans, K. L., Fishburn, I. S. & Gaston, K. J. (2007). Low functional diversity and no redundancy in British avian assemblages. *Journal of Animal Ecology* **76**, 977–985.
Petchey, O. L. & Gaston, K. J. (2002). Functional diversity (FD), species richness and community composition. *Ecology Letters* **5**, 402–411.
Phillips, J. (1931). The biotic community. *Journal of Ecology* **19**, 1–24.
Pielou, E. C. (1972). 2k contingency tables in ecology. *Journal of Theoretical Biology* **34**, 337–330.
R Development Core Team (2010). R: A Language and Environment for Statistical Computing. [2.11.1]. R Foundation for Statistical Computing, Vienna.
Reitalu, T., Prentice, H. C., Sykes, M. T., Lonn, M., Johansson, L. J. & Hall, K. (2008). Plant species segregation on different spatial scales in semi-natural grasslands. *Journal of Vegetation Science* **19**, 407–416.
Ricklefs, R. (1987). Community diversity: relative roles of local and regional processes. *Science* **235**, 167–171.
Ricklefs, R. E. (2008). Disintegration of the ecological community. *American Naturalist* **172**, 741–750.
Roscher, C., Schmid, B. & Schulze, E. D. (2009). Non-random recruitment of invader species in experimental grasslands. *Oikos* **118**, 1524–1540.
Sargent, R. D. & Ackerly, D. D. (2008). Plant-pollinator interactions and the assembly of plant communities. *Trends in Ecology & Evolution* **23**, 123–130.
Schoener, T. W. (1986). Overview: kinds of ecological communities - ecology becomes pluralistic. *Community ecology* (eds J. Diamond and T. J. Case), pp. 467–478. Harper & Row, New York.
Stoll, P. & Weiner, J. (2000). View of interactions among individual plants. *The geometry of interactions: Simplifying spatial complexity* (eds R. Law and J. H. J. Metz), pp. 48–64. Cambridge University Press, Cambridge.
Stone, L. & Roberts, A. (1990). The checkerboard score and species distributions. *Oecologia* **85**, 74–79.
Stubbs, W. J. & Wilson, J. B. (2004). Evidence for limiting similarity in a sand dune community. *Journal of Ecology* **92**, 557–567.
Swenson, N. G. & Enquist, B. J. (2009). Opposing assembly mechanisms in a Neotropical dry forest: implications for phylogenetic and functional community ecology. *Ecology* **90**, 2161–2170.
Turnbull, L. A., Rahm, S., Baudois, O., Eichenberger-Glinz, S., Wacker, L. & Schmid, B. (2005). Experimental invasion by legumes reveals non-random assembly rules in grassland communities. *Journal of Ecology* **93**, 1062–1070.
Ulrich, W. (2004). Species co-occurrences and neutral models: reassessing J. M. Diamond's assembly rules. *Oikos* **107**, 603–609.
Ulrich, W., Almeida, M. & Gotelli, N. J. (2009). A consumer's guide to nestedness analysis. *Oikos* **118**, 3–17.
Vergnon, R., Dulvy, N. K. & Freckleton, R. P. (2009). Niches versus neutrality: uncovering the drivers of diversity in a species-rich community. *Ecology Letters* **12**, 1079–1090.
Wang, P., Stieglitz, T., Zhou, D. W. & Cahill, J. F. (2010). Are competitive effect and response two sides of the same coin, or fundamentally different? *Functional Ecology* **24**, 196–207
Watkins, A. J. & Wilson, J. B. (1992). Fine-scale community structure of lawns. *Journal of Ecology* **80**, 15–24.
Watkins, A. J. & Wilson, J. B. (2003). Local texture convergence: a new approach to seeking assembly rules. *Oikos* **102**, 525–532.
Weiher, E. & Keddy, P. A. (1995). Assembly rules, null models, and trait dispersion: new question from old patterns. *Oikos* **74**, 159–164.
Wiens, J. A. (1989). Spatial scaling in ecology. *Functional Ecology* **3**, 385–397.
Wilson, J. B. (1989). A null model of guild proportionality, applied to stratification of a New Zealand temperate rain forest. *Oecologia* **80**, 263–267.
Wilson, J. B. (1999*a*). Assembly rules in plant communities. *Ecological Assembly Rules: Perspectives, Advances, Retreats* (eds E. Weiher and P. A. Keddy), pp. 130–164. Cambridge University Press, Cambridge.
Wilson, J. B. (1999*b*). Guilds, functional types and ecological groups. *Oikos* **86**, 507–522.
Wilson, J. B. (2007). Trait-divergence assembly rules have been demonstrated: Limiting similarity lives! A reply to Grime. *Journal of Vegetation Science* **18**, 451–452.
Wilson, J. B. & Gitay, H. (1995). Community structure and assembly rules in a dune slack - variance in richness, guild proportionality, biomass constancy and dominance/diversity relations. *Vegetatio* **116**, 93–106.
Wilson, J. B., Gitay, H. & Agnew, A.D.Q. (1987). Does niche limitation exist? *Functional Ecology* **1**, 391–397.

Wilson, J. B., Steel, J. B., Newman, J. E., & King, W. M. (2000). Quantitative aspects of community structure examined in a semi-arid grassland. *Journal of Ecology* **88**, 749–756.

Wilson, S. D. (2000). Heterogeneity, diversity and scale in plant communities. *Ecological consequences of habitat heterogeneity* (eds M. J. Hutchings, E. A. John and A. J. Stewart), pp. 53–69. Blackwell Science, Oxford.

Zhang, J., Hao, Z. Q., Song, B., Li, B. H., Wang, X. G. & Ye, J. (2009). Fine-scale species co-occurrence patterns in an old-growth temperate forest. *Forest Ecology and Management* **257**, 2115–2120.

Zobel, K. & Zobel, M. (1988). A new null hypothesis for measuring the degree of plant community organization. *Vegetatio* **75**, 17–25.

Zobel, K., Zobel, M. & Peet, R. (1993). Change in pattern diversity during secondary succession in Estonian forests. *Journal of Vegetation Science* **4**, 489–498.

Zobel, M. (1997). The relative role of species pools in determining plant species richness: an alternative explanation of species coexistence? *Trends in Ecology & Evolution* **12**, 266–269.

Zobel, M. & Kalamees, R. (2005). Diversity and dispersal—can the link be approached experimentally? *Folia Geobotanica* **40**, 3–12.

Zobel, M. & Partel, M. (2008). What determines the relationship between plant diversity and habitat productivity? *Global Ecology and Biogeography* **17**, 679–684.

VII. SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.

Appendix S1. Studies reviewed herein with information regarding habitat, plant types, approach (according to Table 1), guild and traits, grain size and patch model used within the studies.

Appendix S2. References cited in Appendix S1.

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.

(*Received 11 October 2010; revised 25 May 2011; accepted 26 May 2011; published online 21 June 2011*)

ANNEX 5

A better understanding of the ecological conditions for *Leontopodium alpinum* Cassini in the Swiss Alps

A better understanding of the ecological conditions for *Leontopodium alpinum* Cassini in the Swiss Alps

Mélanie Ischer[1], Anne Dubuis[1], Roland Keller[2] & Pascal Vittoz[1*]

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

[2]Primerose 6, CH-1007 Lausanne, Switzerland.

* Corresponding author:

Pascal Vittoz, Department of Ecology and Evolution, University of Lausanne, Biophore Building, CH-1015 Lausanne, Switzerland; phone: ++41 21 692 4270; fax: ++41 21 692 4165; e-mail: pascal.vittoz@unil.ch

This paper is submitted to Folia Geobotanica.

ABSTRACT

Descriptions of the autecology of *Leontopodium alpinum* are mostly based on empirical knowledge, although this species is considered to be threatened in many countries. In this study, we aimed to define the optimal ecological conditions, in particular, the most important factors, for the presence of *L. alpinum* in the Swiss Alps. The important ecological factors were assessed at the national scale using species distribution models based on topoclimatic predictors and at the community scale using exhaustive plant inventories. The latter were analysed using hierarchical clustering and principal component analysis, and the results were interpreted using ecological indicator values.

L. alpinum was found almost exclusively on base-rich bedrocks (limestone and ultramaphic rocks). The species distribution models showed that the available moisture (dry regions, mostly in the Inner Alps), elevation (mostly above 2000 m.a.s.l.) and slope (mostly >30°) were the most important predictors. The relevés showed that *L. alpinum* is present in a wide range of plant communities, all subalpine-alpine open grasslands, with a low grass cover. As a light-demanding and short-growing species, *L. alpinum* requires light at ground level; hence, it can only grow in open grasslands, with low productivity. These conditions are met in dry conditions (dry, summer-warm climate, rocky and draining soil, south-facing aspect and/or steep slope), at high elevations, on oligotrophic soils and/or on windy ridges. Base-rich soils appear to also be essential. However, years of over-collection have mainly restricted the populations to sites that are not easily accessible.

Keywords alpine grasslands; autecology; phytosociology; species distribution models; Switzerland

INTRODUCTION

Leontopodium alpinum is a perennial herbaceous hemicryptophyte that grows 8-20 cm high (Aeschimann et al. 2004) and is characterised by small yellow capitula surrounded by white and woolly bracts visited by a wide range of insects belonging to 29 families (Erhardt 1993). Its colour, shape, rarity, and legendary inaccessibility have conferred upon it a high symbolic value in alpine regions. Indeed, this species is prized by tourists and botanists (Erhardt 1993); its common name, edelweiss, comes from German and means noble (edel) and white (weiss) (Dweck 2004).

L. alpinum is mostly found in alpine areas, ranging from the Pyrenees to the Central Balkans in Bulgaria (Wagenitz 1979). The genus is native to the Tibetan Plateau and might have migrated during the Pleistocene, when a continual distribution between Asia and Europe was possible (Blöch et al. 2010). This species has been mostly studied with regard to its pharmaceutical properties. It is known for its antibacterial (Dobner et al. 2003) and anti-inflammatory (Dobner et al. 2004) properties and for its analgesic effects (Speroni et al. 2006). In Switzerland, *L. alpinum* has been domesticated (*var. Helvetica*) and is now cultivated by mountain farmers to produce anti-aging creams, sunscreens, and liquor (Carron et al. 2007) or for ornamental purposes (Sigg 2008).

Despite the symbolic value of *L. alpinum* in the Alps, limited scientific information about its autecology is available, with the literature being mostly restricted to descriptive floras based on expert knowledge, and we could not find any descriptive study clearly based on detailed field data. Wagenitz (1979) and Oberdorfer and Müller (1990) indicate that this plant is a light-demanding species found on sunny, rocky grasslands or on cliffs (ledges or crevices) from 1600 m to 2350 m asl (-3000 m) in summer-warm regions on base-rich, mostly calcareous, neutral and mainly humus-rich soils. According to Delarze and Gonseth (2008), this species is found in two distinct phytosociological alliances in Switzerland, *Seslerion* and *Elynion*, both of which are characterised by alkaline to neutral calcareous soils. *Seslerion* is distributed on sunny slopes between 1000 m and 2400 m asl, on dry and stony soils. *Elynion* is restricted to windy ridges between 2000 m and 3000 m, which are only partly protected by snow in winter. Wagenitz (1979) and Oberdorfer and Müller (1990) add *Potentillion caulescentis* on limestone cliffs as a possible alliance. Rey et al. (2011) recently published a synthesis based on previous articles and the extensive experience of the main author, adding to the previous descriptions a preference for a subcontinental climate on slightly dry, oligotrophic soils, with a pH range of 5.5-8.

Although *L. alpinum* is considered as of "least concern" on the Swiss Red List (Moser et al. 2002), it is not a widespread species. Some populations include hundreds of plants;

however, most of the populations are restricted to a few individuals, and the species is far from being present in all *Seslerion* or *Elynion* areas in the Alps. Unfortunately, no monitoring of its populations exists, though many botanists have an impression of decreasing populations. Indeed, this species is considered to be endangered and is protected in most countries or regions where it occurs (cf. review of its status in Rey et al. (2011) and for Switzerland at http://www.infoflora.ch/). Collection by tourists because of the beauty and symbolic value of this species is most likely responsible for the population decreases (Jean 1947; Wagenitz 1979; Rey et al. 2011).

The previous descriptions of *L. alpinum* autecology are certainly trustworthy, yet, to ensure the optimal conservation of a species, it is important to extend beyond expert knowledge by adding analyses based on precise observations and by identifying the dominant ecological factors to explain its distribution. Distribution models are efficient tools to obtain reliable information on the ecology of species (Guisan and Zimmermann 2000) and produce faithful habitat suitability maps (Le Lay et al. 2010), and the use of such models is possible with rare or uncommon species (Engler et al. 2004). Exhaustive phytosociological relevés, which are numerous in Switzerland (Schaminée et al. 2009), offer a quick way to obtain data on many populations, with partner species giving indirect indications of ecological conditions (Deil 2005; Vittoz et al. 2006).

This study aimed to improve our knowledge regarding the necessary ecological conditions for *L. alpinum* in the Swiss Alps. In particular, we aimed to identify the most important ecological factors explaining the distribution of this species and to obtain the most complete overview of the habitats in which the plant can grow. For this, we investigated the existing data with (1) species distribution models to identify important ecological factors at the national scale and produce a map of potential habitats and (2) clustering and multivariate analyses of exhaustive plant inventories, interpreted with the help of ecological indicator values, to study the species ecological range at the community scale.

MATERIAL AND METHODS

Study area

The entire area of the Swiss Alps, which represents 60% of the Swiss territory (App. 1 in supplemental archives), was considered in this study. Due to their proximity to the Atlantic Ocean and the Mediterranean Sea, the Outer Alps are characterised by a wet, suboceanic climate. Conversely, some valleys in the Inner Alps are protected from rainfall by high mountains and experience a dry, subcontinental climate.

Floristic data

The existing data on *L. alpinum* were collected, either in the form of exhaustive species lists (phytosociological relevés) or from isolated observations. Relevés were collected from the literature (Braun-Blanquet 1969; Galland 1982; Reinalter 2004; Steiner 2002), from personal data obtained in previous projects (Randin et al. 2010) and from unpublished data (M. Ischer, J.-L. Richard, M. Schütz, R. Keller). All Swiss relevés were retained, without consideration of the size of the inventoried plot or the availability of exact coordinates. A total of 249 relevés were collected. The vascular plant nomenclature is according to Aeschimann et al. (2004).

Isolated observations were mainly provided by Info Flora (Swiss Floristic Network; http://www.infoflora.ch/) and completed by occurrences in relevés. Only observations with a horizontal accuracy of 100 metres or less were retained for the models, corresponding to a total of 344 occurrences.

Environmental data

The species distribution in relation to geology was evaluated by counting the number of *L. alpinum* occurrences in each geological category. We used a geotechnical map at a 1:200,000 scale (Swiss Geotechnical Commission, http://www.sgtk.ch) in which the substrates are divided into 30 categories defined by bedrock type or granulometry for recent deposits. These categories were simplified into four categories: purely calcareous bedrocks, mixed bedrocks potentially with limestone (e.g., moraines, alluvial deposits, conglomerates), ultramaphic bedrocks and base-poor bedrocks (e.g., granite, quartzite).

In the species distribution models, we used three climatic and two topographic predictors, all generally considered as important and complementary ecological variables to explain species distribution in mountain environments (Körner 2003). The climatic predictors were calculated based on temperature and precipitation data recorded by MeteoSwiss (www.MeteoSwiss.ch) and interpolated with a 25-m resolution digital elevation model (see

Zimmermann and Kienast 1999 for methodology). We used the mean temperature for the growing season (June to August, in °C), the average moisture index over the growing season (average value of the balance between precipitation and potential evapotranspiration in $mm \cdot day^{-1}$) and the sum of solar radiations for the growing season (in $kJ \cdot m^{-2} \cdot yr^{-1}$). The topographic predictors were derived from the digital elevation model. We used the slope (in degrees) and the topographic position (an integrated measure of topographic features; Zimmermann et al. 2007); positive values of topographic position indicate ridges, whereas negative values indicate valleys.

Species distribution models

To restrain the study area considered in the models to plausible areas for *L. alpinum*, a mask was created by removing the urbanised areas, glaciers and lakes. Moreover, only elevations higher than 1300 m asl were retained because no current occurrence is reported at lower elevations and the large majority of observations are above 1500 m.

Out of the 344 available observations, we randomly selected occurrences separated by a minimum distance of 250 metres to avoid spatial autocorrelation. This selection was repeated 20 times, retaining an average of 214 occurrences. As no absences were available, we used pseudo-absences to calibrate the models (Engler et al. 2004). Using Hawth's Analysis Tools in ArcGIS (2004), we randomly generated 10'000 pseudo-absences in the entire study area, as defined in the preceding paragraph. We used all the pseudo-absences during the model calibration but gave to each of them a weight equal to the number of occurrences divided by 10'000, as recommended by Barbet-Massin et al. (2012).

The models were computed using the library BioMod (Thuiller et al. 2009) in R (v.2.14.1; R Development Team 2011). Four model types were computed to obtain a reliable probability regarding the presence of *L. alpinum*: generalised linear models (GLM; McCullagh and Nelder 1989) with a polynomial term and a stepwise procedure (computed with the Akaike information criteria, AIC), generalised additive models (GAM; Hastie and Tibshirani 1990) using a spline function with a degree of smoothing of 4, generalised boosting models (GBM; Ridgeway 1999; Friedman et al. 2000) with 2000 trees and random forest (RF; Brieman 2001).

As no independent data were available to evaluate the models, we used a split-sample procedure and 70% of the data (presences and pseudo-absences) were randomly chosen and used for model calibration, with 30% used for model evaluation. This procedure was repeated 5 times for each model type and for each of the 20 datasets of occurrence selection. For each data split, the predictive performance of the models was evaluated with

two frequently used metrics: the area under the curve (AUC) of a receiver operating characteristic plot (ROC; Ogilivie and Creelman 1968) and the Boyce index (Boyce et al. 2002). The AUC varies from 0.5 for models providing random predictions to 1 for perfect models; a model is considered reliable if its AUC is higher than 0.7 (Swets 1988). The Boyce index varies between -1 for counter predictions and 1 for perfectly reliable predictions; values close to 0 indicate random predictions. This last index is particularly adapted here as it uses only presences (Hirzel et al. 2006).

As proposed in BioMod, the relative importance of each predictor in a model was calculated by randomising one predictor and recalibrating the model with the randomised variable while keeping the other predictors unchanged. The results of the model containing the randomised predictor were then correlated with those of the original model. The importance of the predictor was calculated as one minus this correlation (consequently, the higher the value is, the more important is the variable for the model quality). The calculation of predictor importance was repeated five times for each model and for each of the 20 datasets of occurrence selection.

To produce the most reliable map of potential habitat, we used an ensemble forecasting, as recommended by Araujo and New (2007), with the final set of models calibrated using 214 randomly selected occurrences and the pseudo-absences. The results of the four models were average weighted by their respective AUC. The probabilities of presence predicted by the ensemble model were transformed into presence-absence data with an optimised threshold maximising the model sensitivity and specificity (Liu et al. 2005). The projection of the ensemble model was finally restricted to the three geological categories on which *L. alpinum* was most commonly observed (see the Results section).

Analyses of relevés

All the subsequent analyses of the 249 relevés were realised following the elimination of rare species (< 3 occurrences) and transformation of the cover indices (Braun-Blanquet 1964), as follows: r=1, +=2, 1=3, 2=4, 3=5, 4=6 and 5=7.

To distinguish groups of phytosociological relevés, a hierarchical clustering analysis was performed using the chord distance and Ward's minimum variance clustering method (Borcard et al. 2011). The optimal number of groups was defined according to Mantel correlation coefficients between the distance matrix and binary matrices (0 if relevés are in the same group, 1 otherwise) computed from the dendrogram sectioned at various levels (see the function in R given by Borcard et al. 2011, p. 71). Differential species in the groups were selected by calculating their indicator values (Dufrêne and Legendre 1997).

The ecological conditions of each group were documented by the eight ecological indicator values attributed to species by Landolt et al. (2010). The median value was calculated for each relevé and each indicator value. However, to obtain more informative values than the habitual median limited to the indicator classes, the median μ was calculated with the following formula based on the 50% value of a cumulative frequency curve (Ellenberg 1991):

$$\mu = m + w\frac{(n/2) - n_m}{n_x}$$

where m is the lower limit of the median class (here, the intermediate value between the median class and the previous one), w is the width of the classes (0.5 for T and F; 1 for L, K, R and N; 2 for D and H; see Table 2 for abbreviations), n is the number of species in the relevé, n_m is the number of species with an indicator value lower than the median class and n_x is the number of species with an indicator value similar to the median class.

A principal component analysis (PCA), after a Hellinger transformation of the data (Legendre and Gallagher 2001), was computed to study the distribution of the groups along floristic gradients. The ecological indicator values (Landolt et al. 2010) were passively projected using the correlation between the median ecological values of each relevé and its scores on axes (Wohlgemuth 2000).

These analyses were computed using R software (R Development Team 2011) with the libraries *vegan* (clustering, Hellinger transformation, PCA) and *labdsv* (indicator values of the species). The nomenclature of the phytosociological alliances follows Delarze & Gonseth (2008), and the names of associations were conserved from the authors of the relevés (not all the relevés were classified).

RESULTS

Out of the 344 available occurrences, 242 were on pure calcareous bedrocks, 49 on mixed bedrocks, 42 on ultramaphic bedrocks and 11 on base-poor bedrocks.

Species distribution models

The evaluation values obtained for the four modelling techniques and the 20 datasets (random selection of presences) after the split-sample procedure resulted in a mean AUC value of 0.81±0.01 and a mean Boyce index of 0.90±0.18. The four modelling techniques and the 20 datasets produced predictions of similar quality.

The most important predictor for modelling the *L. alpinum* distribution was the average moisture index (Fig. 1), followed by the summer temperature, the slope and the topographic position. Solar radiation was the least important predictor. The response curves of the models showed that the optimal value for the moisture index was less than 5 $mm \cdot day^{-1}$, which corresponds to dry conditions for the Alps (Fig. 2). The suitability of habitat increased with a decreasing mean temperature (i.e., with increasing elevation) and increasing slope.

The map of predicted potential habitats showed that the Inner Alps are the most suitable area for *L. alpinum*, with isolated possible regions in the Outer Alps (App. 1 in supplemental archives).

Clustering and principal component analysis of the relevés

The first axis of the PCA explained 10.1% of the variance, and the ecological indicator values pointed to a temperature and light gradient along this axis (Fig. 3), extending from a heliophilous pole on the left side to a thermophilic pole on the right side. The second axis explained 7.4% of the variance and was associated with a gradient of soil pH.

Eight relevé groups were retained in the clustering analysis (Table 1). The median ecological indicator values (Landolt et al. 2010; Table 2) showed high ranges between the groups for soil pH (3.36-4.67, neutral to alkaline), soil aeration (3.16-4.40, moderate to good aeration), temperature (1.33-2.25, lower subalpine to alpine belts) and humus content (2.00-2.92, little to moderate). Moderate ranges of values were observed for the light conditions (4.09-4.81, well-lit sites to almost full light), whereas small ranges were found for humidity (moderately dry to fresh), continentality (subcontinental) and nutrient availability (infertile).

Group 3 contained the most heliophilous relevés under the coldest conditions and were situated on humus-poor soils but with good aeration (Table 2). This group was differentiated by scree and rock species (*Saxifraga oppositifolia, Carex rupestris, Herniaria alpina*).

Previous classifications of these relevés were mainly attributed to *Herniarietum alpinae* Zollitsch 1968 found on screes of calcareous slates (*Drabion hoppeanae*). Close to this group, group 2 shared *Elyna myosuroides* (highest frequency in this group) with two other *Elynion* species (*Ligusticum mutellinoides*, *Arenaria ciliata*). This group had the most humus-rich soils. The previously classified relevés were mostly attributed to *Elynetum myosuroidis* Rübel 1911 (*Elynion*), corresponding to alpine windy ridges.

At the other extreme of the light-temperature gradient, group 7 corresponded to the warmest and least lit (densest grass cover; Table 2) conditions and was differentiated by species belonging to continental steppes (*Stipo-Poion: Carex humilis, Koeleria macrantha, Pulsatilla halleri*, *Erysimum rhaeticum*) and other species from dry, thermophilous grasslands (e.g., *Euphorbia cyparissias, Teucrium montanum*, *Dianthus sylvestris*, *Plantago serpentina*, *Galium lucidum*). Almost all of these relevés are from the Zermatt region, in the Inner Alps, and were attributed to *Astragalo leontini-Seslerietum* Richard 1985, which is an association representing the dry, continental wing of *Seslerion* (Steiner 2002). Close to this group, group 1 is similarly characterised by poorly lit conditions and moderate soil aeration but colder conditions. All of the differential species (*Senecio doronicum, Carex sempervirens, Anthyllis vulneraria subsp. Alpestris*, *Festuca violacea aggr.*) are typical of calcareous, alpine grasslands (*Seslerion*). The previously classified relevés were attributed to *Seslerio-Caricetum sempervirentis* Br.-Bl. in Br.-Bl. et Jenny 1926, the central association of the *Seslerion*.

The largest group (6) showed the lowest soil pH (R, Table 2), corresponding to a mixture of calcicolous species (e.g., *Draba aizoides*, *Oxytropis helvetica*) and species colonising neutral to weakly acidic soils (e.g., *Artemisia glacialis*, *Veronica fruticans*). *Carex curvula s.l.* is problematic, as it contains an acidophilous taxon (*C. curvula s.str.*) and a calcicolous taxon (*C. curvula subsp. rosae*), though the exact taxon was not always indicated. However, 38 of the 39 precisely identified occurrences were *C. curvula subsp. rosae*. Three-quarters of the relevés were previously attributed to *Artemisio glacialis-Festucetum pumilae* Richard 1985, a pioneer association of *Seslerion* on little-developed soils, rich in gravel and sand, partly unstable, in the Inner Alps (Steiner 2002), whereas one-quarter was attributed to *Elynetum myosuroidis*.

Groups 4 and 5 presented intermediate compositions, without any frequent differential species and showing median ecological values mostly around the middle of the ranges (Table 2). The relevés were previously attributed to various associations: *Artemisio glacialis-Festucetum pumilae*, *Seslerio-Caricetum sempervirentis*, *Androsacetum alpinae* Br.-Bl. 1918 (fine screes of siliceous or ultramafic rocks, in *Androsacion alpinae*), *Caricetum*

fimbriatae Richard 1985 (screes of ultramaphic rocks, in *Caricion curvulae*), *Androsacetum helveticae* Br.-Bl. in Br.-Bl. et Jenny 1926 (alpine calcareous cliffs, in *Potentillion caulescentis*) or *Potentillo caulescentis-Hieracietum humilis* Braun-Blanquet 1933 (montane-subalpine calcareous cliffs, in *Potentillion caulescentis*). All these associations are linked to rocky conditions, with a low vegetation cover.

Lastly, group 8 was separated along the second axis of the PCA (Fig. 3) and was characterised by the highest soil pH (Table 2) and was differentiated by six species found on rocky, open grasslands on limestone (*Dryas octopetala*, *Saxifraga caesia*, *Carex mucronata*, *Gentiana clusii*, *Carex firma*, *Crepis kerneri*). All of the previously classified relevés were attributed to *Caricetum firmae* Rübel 1911, with rocky, calcareous grasslands in the alpine belt (*Caricion firmae*).

DISCUSSION

In this study, we investigated the autecology of *Leontopodium alpinum* to obtain a better understanding of its distribution in the Swiss Alps. Although it is an emblematic species, very few published data regarding its ecology are based on field measurements. We used species distribution models to assess its distribution at the national scale in relation to topoclimatic factors and exhaustive plant inventories (phytosociological relevés) to encompass the entire ecological range of the plant communities in which the species can be found.

As our species distribution models obtained predictions with AUC values above 0.8 and Boyce index values close to 1, they can be qualified as good and trustworthy (Araujo et al. 2005). Based on the model results, *L. alpinum* grows in dry areas, at high elevations and on steep slopes.

The collected relevés were realised by different authors within the context of vegetation studies, mainly aiming to classify plant communities. Hence, we can consider them not to be biased toward particular conditions linked to *L. alpinum* growth. However, there is a regional bias, as approximately half of the 249 available relevés were from the Zermatt region, which is a particularly interesting region for flora and plant communities and has been extensively investigated (Richard 1991; Steiner 2002). Nevertheless, the other available relevés are well distributed throughout the Swiss Alps, including the Outer Alps; altogether, they most likely represent most of the ecological range of the species in the Swiss Alps. *L. alpinum* was observed in a broad range of phytosociological associations, mainly belonging to grasslands of the alliances *Seslerion, Elynion* or *Caricion firmae*, with some occurrences in cliffs of *Potentillion caulescentis* and calcareous screes of *Drabion hoppeanae* or in associations restricted to screes of ultramaphic rocks belonging to *Androsacion alpinae* and *Caricion curvulae*. These plant communities are all subalpine-alpine open grasslands on base-rich bedrocks, with a low to very low grass cover, on nutrient-poor, dry to fresh soils, among apparent rocks or on ridges (Delarze & Gonseth 2008; Steiner 2002)

Ecological conditions for *L. alpinum*

As indicated in previous descriptions (Wagenitz 1979; Oberdorfer and Müller 1990), *L. alpinum* was mostly observed on limestone, with 85% of the occurrences on pure calcareous bedrocks or on mixtures containing limestone (e.g., moraines). However, other basic cations can replace calcium, as 12% of the observations were on ultramaphic bedrocks, previously only mentioned by Wagenitz (1979). Conversely, the species avoids

siliceous bedrocks, with only 3% of the occurrences, whereas this type of rocks represents 39% of the Swiss Alps above 1300 m asl. As contamination by closely located limestone cannot be excluded for some of these rare occurrences, the proportion of occurrences on siliceous rocks is certainly lower. The high median ecological indicator value for soil pH (R) for all the relevé groups (Table 2) confirmed the geological observations. However, this indicator also showed the highest variation among the groups, extending from weakly acid to alkaline soils. This corresponds to the pH of 5.5-8 reported by Rey et al. (2011) and indicates that the soil is consistently base-rich but can be completely decarbonated.

Another constant factor for all the relevés is the low grass cover, with only communities of open grasslands. *L. alpinum* is a light-demanding species (Wagenitz 1979), with most of the leaves in a rosette on the ground, and it most likely does not tolerate competition by other species. The importance of light at the soil level is shown by the constant species (Table 1; *Sesleria caerulea, Festuca quadriflora, Minuartia verna*, *Aster alpinus* and *Agrostis alpina*), which are all heliophilous (Landolt et al. 2010), and by the median ecological value for light above 4 (well lit to full light) for all the relevé groups. This low grass cover is provided by rocky conditions, sometimes in pioneer communities on slightly unstable screes, or by other harsh conditions limiting plant growth (see below). Although the species is indicated as growing in sunny conditions (Oberdorfer and Müller 1990), solar radiation was not important as a predictor in the models, which most likely means that the aspect alone is not a constraint, with some stands occurring on a north aspect, and that a south-facing slope may be a way to limit plant growth through dry conditions.

Indeed, the most important predictor in the models was a low moisture index, translating into a preference for regions with an approximate balance between rainfall and potential evapotranspiration or dryer. In Switzerland, this corresponds to the subcontinental climate of the Inner Alps, as indicated by Rey et al. (2011), though the species is present in wetter regions as well. This finding is in agreement with the median ecological indicator values for continentality (K, subcontinental climate) and for soil humidity (F), ranging between moderately dry to fresh. Previously, Wagenitz (1979) characterised soil humidity for *L. alpinum* as relatively dry. The moderate to good aeration of the soil (indicator value D) can be considered to be a contribution to the good water drainage.

Temperature was the second most important predictor in the models, with a higher suitability for mean summer temperatures below 10°C, corresponding approximately to elevations >2000 m asl. This preference for high elevations has long been clear, though some isolated observations in lowlands have been reported (e.g., 220 m asl in Slovenia and 470 m in a Swiss wetland; Wagenitz 1979; Rey et al. 2011). Moreover, the possibility of

cultivating *L. alpinum* in montane belt and its ability to grow with highly thermophilous species in group 7 (e.g., *Galium lucidum, Astragalus monspessulanus*, *Phleum phleoides* have their optimum in the warm colline belt; Landolt et al. 2010) could indicate that low temperatures are not a necessary condition but indirectly helpful for limiting competition.

L. alpinum has the reputation to grow in sites that are not easily accessible, on cliffs or steep slopes, corresponding to the steep slope (mainly >20°) retained in the models. However, slope had a low importance, and many populations are situated on flat lands. Wagenitz (1979) and Rey et al. (2011) indicated that the observed inaccessibility corresponds to the populations remaining after decades of over-collection for the tourism market. However, cliffs and steep slopes are most likely also helpful to restrict competition by providing dryer conditions and regular erosion.

Overall, the projection of the ensemble model predicted the potential habitat of *L. alpinum* to mainly fall within the Inner Alps, with only isolated potentially favourable regions occurring in the Northern and Southern Outer Alps, though limestone is by far the dominant bedrock in the Northern Outer Alps. As shown in the community analysis, the most important factor for the growth of this plant is certainly an abundance of light at the ground level, indicating a grassland with low productivity. Growth under subcontinental, dry conditions (Rey et al. 2011), with summer-warm temperatures (Wagenitz 1979), is most likely an efficient way for *L. alpinum* to limit competition with taller species, whereas the Outer Alps, with their abundant rainfalls and denser grasslands, are less favourable. Oligotrophic soils (all groups with a low ecological value for N), high elevations, dry southern aspect, steep slopes, raw soils or windy ridges, where species have to withstand very cold conditions because of the absence of snow in winter (Vonlanthen et al. 2006), are other complementary or substitute stressful conditions that limit competition. However, the importance of base-rich bedrocks and soils is less clear and could be interpreted as a supplementary factor reducing plant growth and, hence, competition with other species because of the limited availability of many essential cations (Duchaufour 1995) and the often strong drainage on limestone. However, base-rich conditions are most likely physiologically necessary for *L. alpinum*. Indeed, Wagenitz (1979) stated that the species was never found on strong acidic silicate bedrocks, and none of the available relevés were from siliceous cliffs or other acidophilous, dry grasslands, though some are open communities. The physiological relationship ought to be investigated in future studies.

Conclusions and perspectives

This study allowed us to obtain a precise description of the ecological requirements of *Leontopodium alpinum*, mostly confirming the previous, empirical descriptions of its autecology yet helping to prioritise the different ecological factors. The two essential factors are a considerable amount of light at the ground level and a base-rich soil. As a short-statured, light-demanding species, *L. alpinum* does not withstand competition from other species. All the other ecological characteristics can be interpreted as ways to limit competition by stressful conditions (e.g., high elevations, windy ridges, southern aspect, steep slopes, oligotrophic soils, rocks and cliffs, draining substrates). The different possible combinations of these conditions result in a broad range of plant communities in which *L. alpinum* can grow. Some of them, such as screes of the *Drabion hoppeanae* and *Androsacion alpinae*, where not mentioned previously in the literature.

The projection of the models pointed to many potential areas for *L. alpinum* in the Swiss Alps. However, based on our experience, we know that not all of these areas are colonised. Supplementary investigations are, therefore, necessary to evaluate why the species is so scattered and uncommon. Previous collection by tourists is certainly an important cause, but biogeographical reasons linked to its historical distribution, recent developments in the Alps, such as the marked increase of sheep herds (FSO 2010), and possible recruitment limitations due to poor seed production, dispersal capacities or establishment rate need to be addressed as potential causes to explain local absences. A limited ability to disperse (Handel-Mazzetti 1927) could be an important problem in the middle to long term when considering the scattered distribution of this species and future climate change (IPCC 2007).

ACKNOWLEDGEMENTS

We thank Info Flora and Martin Schütz from the Forschungsanstalt für Wald, Schnee und Landschaft (WSL) for providing data and "Fondation Dr Ignace Mariétan" and Weleda for partial funding. We are very grateful to Antoine Guisan for helpful advice regarding the modelling techniques and Tomas Herben and two anonymous reviewers for their useful comments on an earlier draft of the manuscript.

REFERENCES

Aeschimann D, Lauber K, Moser DM, Theurillat J-P (2004) *Flora alpina*. Belin, Paris

Araujo MB, Pearson RG, Thuiller W, Erhard M (2005) Validation of species–climate impact models under climate change. *Global Change Biol* 11:1504–1513

Araujo MB, New M (2007) Ensemble forecasting of species distributions. *TRENDS Ecol Evol* 22:42-47

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol Evol* 3:327–338

Blöch C, Dickoré WB, Samuel R, Stuessy TF (2010) Molecular phylogeny of the Edelweiss (Leontopodium, Asteraceae – Gnaphalieae). *Edinburgh J Bot* 67:235-264

Borcard D, Gillet F, Legendre P (2011) *Numerical ecology with R*. Springer, New York

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. *Ecol Model* 157:281–300

Braun-Blanquet J (1964) *Pflanzensoziologie. Grundzüge der Vegetationskunde*. Ed. 3, Springer, Wien - New-York

Braun-Blanquet J (1969) *Die Pflanzengesellschaften der rätischen Alpen im Rahmen ihrer Gesamtverbreitung. I. Teil*. Bischofberger & Co, Chur

Brieman L (2001) Random Forests. *Mach learn* 45:5-32

Carron C-A, Previdoli S, Baroffio C (2007) Helvetia, une nouvelle variété d'edelweiss issue d'hybrides de clones. *Rev suisse Vitic Arboric Hortic* 39:125-130

Deil U (2005) A review on habitats, plant traits and vegetation of ephemeral wetlands - a global perspective. *Phytocoenologia* 35:533-706

Delarze R, Gonseth Y (2008) *Guide des milieux naturels de Suisse. Ecologie, menaces, espèces caractéristiques*. Rossolis, Bussigny

Dobner MJ, Schwaiger S, Jenewein IH, Stuppner H (2003) Antibacterial activity of Leontopodium alpinum (Edelweiss). *J Ethnopharmacol* 89:303-301

Dobner MJ, Sosa S, Schwaiger S, Altinier G, Loggia RD, Kaneider NC, Stuppner H (2004) Anti-inflammatory activity of Leontopodium alpinum and its constituents. *Planta Med* 70:502-508

Duchaufour P (1995) *Pédologie. Sol, végétation, environnement.* Ed 4, Masson, Paris

Dufrêne M, Legendre P (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol Monogr* 67:345-366

Dweck AC (2004) A review of Edelweiss. *SÖFW-Journal* 130:65-68

Ellenberg H (1991) *Zeigerwerte der Gefässpflanzen (ohne Rubus)*. In Ellenberg H, Weber HE, Düll R, Wirth V, Werner W, Paulissen D (eds) Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta Geobotanica* 18. Erich Golze KG, Göttingen, pp 9-166

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. *J Appl Ecol* 41:263-274

Erhardt A (1993) Pollination of edelweiss, Leontopodium alpinum. *Bot J Linn Soc* 111:229-240

Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: A statistical view of boosting - Rejoinder. *Ann Stat* 28:400-407

FSO (2010) *Agriculture Suisse – Statistique de poche 2010*. Swiss Federal Statistical Office, Bern

Galland P (1982) *Etude de la végétation des pelouses alpines au Parc national suisse*. PhD thesis, Université de Neuchâtel, Neuchâtel

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Model* 135:147-186

Handel-Mazzetti H (1927) Systematische Monographie der Gattung Leontopodium. *Beihefte zum Botanischen Centralblatt* 44:1-178

Hastie TJ, Tibshirani R (1990) *Generalized additive models*. Chapman and Hall, London

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. *Ecol Model* 199:142–152

IPCC (2007) *Summary for Policymakers*. In Solomon S et al (eds) 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, Cambridge

Jean L (1947) *Fleurs des Alpes*. Ophrys, Paris

Körner C (2003) *Alpine plant life.* Springer, Berlin

Landolt E, Bäumler B, Erhardt A, Hegg O, Klötzli F, Lämmler W et al (2010) *Flora Indicativa*. Haupt Verlag, Berne

Le Lay G, Engler R, Franc E, Guisan A (2010) Prospective sampling based on model ensembles improves the detection of rare species. *Ecography* 33:1015-1027

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271-280

Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385–393

McCullagh P, Nelder JA (1989) *Generalized Linear Models*. Ed. 2, Chapman & Hall, London

Moser D, Gygax A, Bäumler B, Wyler N, Palese R (2002) *Liste rouge des fougères et plantes à fleurs menacées de Suisse*. Office fédéral de l'environnement, des forêts et du paysage, Centre du Réseau suisse de floristique, Conservatoire et Jardin botanique de la Ville de Genève, Bern, Chambésy

Oberdorfer E, Müller T (1990) *Pflanzensoziologisch Exkursionsflora*. Ulmer, Stuttgart

Ogilivie JC, Creelman CD (1968) Maximum-likelihood estimation of receiver operating characteristic curve parameters. *J Math Psychol* 5:377-391

Randin CF, Engler R, Pearman PB, Vittoz P, Guisan A (2010) *Using georeferenced databases to assess the effect of climate change on alpine plant species and diversity*. In Spehn E, Körner C (eds) *Data mining for global trends in mountain biodiversity*. CRC Press, Taylor & Francis Group, Boca Raton, pp 149-163

Reinalter R (2004) *Zur Flora der Sedimentgebiete im Umkreis der Südrätischen Alpen, Livignasco, Bormiese und Engiadin'Ota (Schweiz-Italien).* Birkhäuser, Basel

R Development Core Team (2011) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna

Rey C, Rey S, Baroffio C, Vouillamoz JF, Roguet D (2011) *Edelweiss, reine des fleurs.* Ed. du Belvédère, Fleurier

Richard J-L (1991) Flore et végétation de Zermatt (VS): premier aperçu et réflexions. *Bull Murith* 109:27-40

Ridgeway G (1999) The state of boosting. *Comput Sci Stat* 31:172-181

Schaminée JHJ, Hennekens SM, Chytry M, Rodwell JS (2009) Vegetation-plot data and databases in Europe: an overview. *Preslia* 81:173-185

Sigg P (2008) Culture de l'edelweiss pour la fleur coupée. *Rev suisse Vitic Arboric Hortic* 40:349-356

Speroni E et al. (2006) In vivo efficacy of different extracts of Edelweiss (Leontopodium alpinum Cass.) in animal models. *J Ethnopharmacol* 105:421-426

Steiner A (2002) Die Vegetation der Gemeinde Zermatt. *Geobotanica Helvetica* 74:1-204

Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* 240:1285-1293

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

Vittoz P, Wyss T, Gobat J-M (2006) Ecological conditions for Saxifraga hirculus in Central Europe: A better understanding for a good protection. *Biol Conserv* 131:594-608

Vonlanthen CM, Bühler A, Veit H, Kammer PM, Eugster W (2006) Alpine plant communities: a statistical assessment of their relation to microclimatological, pedological, geomorphological, and other factors. *Phys Geogr* 27:137-154

Wagenitz G (1979) *Compositae I: Allgemeiner Teil, Eupatorium-Achillea*. In Hegi G, Conert HJ, Hamann U, Schultze-Motel W, Wagenitz G (eds) *Illustrierte Flora von Mitteleuropa. Band VI, Angiospermae Dicotyledones 4, Teil 3*. Ed. 2, Paul Parey, Berlin, pp 133-136

Wohlgemuth T (2000) Diskreter und kontinuierlicher Charakter der Vegetation — Waldvegetationsdaten als Referenz. *Bauhinia* 14:67-88

Zimmermann NE, Edwards TC, Moisen GG, Frescino TS, Blackard JA (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *J Appl Ecol* 44:1057-1067

Zimmermann NE, Kienast F (1999) Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *J Veg Sci* 10:469-482

TABLES

Table 1 Synthetic table of the clustering analysis. Eight groups of relevés were retained, and the species were classified based on their indicator values for the groups. The groups are approximately ordered following axis 1 of the PCA (see Fig. 3). Only species present in at least 40% of the relevés of one group are retained in this table (see App. 2 for the complete table). Species frequency in the groups is given by the Roman numeral: V, species frequency >80%; IV, 60-80%; III, 40-60%; II, 20-40%; I, 10-20%; r, <10%.

Relevés groups	3	2	6	5	1	4	7	8
Nbre of relevés	13	33	63	30	38	30	29	13
Constant species								
Leontopodium alpinum	V	V	V	V	V	V	V	V
Sesleria caerulea	I	V	IV	V	V	II	V	V
Festuca quadriflora	V	V	V	IV	III	III	II	III
Minuartia verna	V	IV	IV	IV	III	II	IV	II
Aster alpinus	I	IV	IV	IV	IV	III	V	II
Agrostis alpina	II	V	IV	IV	III	II	IV	II
Elyna myosuroides	IV	V	II	III	I	I	I	r
Galium pumilum	r	r	IV	II	III	r	III	r
Gentiana verna	I	III	IV	II	II	II	III	r
Thymus serpyllum aggr.	r	II	III	III	IV	IV	V	r
Campanula cochleariifolia	II	r	I	III	II	r	r	I
Differential species group 3								
Saxifraga oppositifolia	V	III	II	II	I		r	
Carex rupestris	V	I	r	IV	r			
Herniaria alpina	IV	r	IV	II		I	II	
Gentiana schleicheri	IV	I	III	II				
Linaria alpina s.str.	III	r	I	r			II	
Draba dubia	III	r	r	I				
Differential species group 2								
Polygonum viviparum	III	V	r	III	III	r		II
Campanula scheuchzeri		IV	III	II	III	I	I	r
Arenaria ciliata	III	III	II	I	r			r
Ligusticum mutellinoides	III	III	II		r	r		
Pedicularis verticillata		III	r	I	I			r
Silene acaulis		III	r	r	r			I
Differential species group 6								
Draba aizoides	IV	II	V	II	I	II	II	
Carex curvula s.l.		II	IV	r		I		
Sempervivum arachnoideum	r	II	IV	II	I	III	III	
Artemisia glacialis			III		r	III		
Veronica fruticans		r	III	r	r	I	II	
Oxytropis helvetica	II		III	r				
Differential species group 1								
Senecio doronicum		r	III	III	V	II	IV	
Carex sempervirens		III	r	III	V	I	I	r
Anthyllis vulneraria subsp. alpestris		II	I	II	IV	I	II	II
Festuca violacea aggr.		r	II	I	III	r	I	
Phyteuma orbiculare		r			III			
Scabiosa lucida		r			III			
Differential species group 7								
Bupleurum ranunculoides s.str.		r	r	II	I	III	V	
Euphorbia cyparissias		r	II	r	II	II	V	
Carex humilis		r		I	II	r	V	
Oxytropis campestris s.str.		II	II	I	II	III	IV	
Dianthus sylvestris			r	r	r	III	IV	
Acinos alpinus		r	I	r	II	II	IV	

Relevés groups	3	2	6	5	1	4	7	8
Nbre of relevés	13	33	63	30	38	30	29	13
Differential species group 7 (continuation)								
Hieracium pilosella			I	I	r	II	IV	
Koeleria macrantha			r			II	IV	
Carlina acaulis subsp. caulescens		r	r	I	II	I	IV	
Teucrium montanum			r	I	r	I	IV	r
Astragalus australis			r	I	I	II	III	
Plantago serpentina			r	r	r	II	III	
Pulsatilla halleri			I	r	r	II	III	
Erysimum rhaeticum			r		I	II	III	
Campanula rotundifolia		r		I	I	I	III	
Trifolium montanum		r		r	r	r	III	
Astragalus leontinus			r		r	I	III	
Briza media				r	I	r	III	
Carex caryophyllea		r			r	r	III	
Galium lucidum				r	r	I	III	
Dactylis glomerata						r	III	
Linum catharticum					I		III	
Differential species group 8								
Dryas octopetala		II	r	I	II			V
Saxifraga caesia		r		I	r			V
Carex mucronata			r	r	r	I	r	IV
Gentiana clusii		I		I	II			IV
Carex firma				I	r			IV
Crepis kerneri				r				III
Other species								
Helianthemum alpestre		II	II	IV	III	II	III	V
Carduus defloratus s.l.		r	I	I	IV	r	IV	r
Galium anisophyllon	r	III	IV	III	II	I	IV	
Euphrasia salisburgensis		I	r	II	III	r	II	I
Sedum atratum	III	I	I	I	II	I		r
Gypsophila repens	I		r	III	II	I	III	r
Euphrasia minima	I	III	II	I	I	I	r	
Helianthemum nummularium s.l.		I	II	II	V	II	V	
Festuca ovina aggr.		I	II	II	II	III	IV	
Lotus corniculatus aggr.		r	I	I	III	I	IV	
Cerastium arvense subsp. strictum		r	III	I	I	III	II	
Gentiana campestris s.str.		II	I	I	III	r	I	
Hieracium villosum		r		r	III	r	II	I
Globularia cordifolia		r		II	II	II	III	II
Juniperus communis subsp. nana		r	I	II	II	I	III	
Silene exscapa	III	II	II	r		I		
Euphrasia alpina		r	II	I		II	III	
Anthyllis vulneraria subsp. valesiaca			III	II	r	I	III	
Leucanthemum adustum				r	III	r	III	
Mean number of species	16.6	29	29.3	24.7	35.1	24.6	40.8	14.6

Table 2 Median ecological indicator values (Landolt et al. 2010) of the relevé groups. The groups are ordered following axis 1 of the PCA (Fig. 3). The highest value of each indicator is in **bold** and the lowest in *italics*.

Relevé group	3	2	6	5	1	4	7	8
Climate								
T (temperature)	*1.33*	1.55	1.57	1.73	1.88	1.89	**2.25**	1.63
L (light)	**4.81**	4.41	4.50	4.41	*4.09*	4.32	*4.09*	4.67
K (continentality)	3.85	3.61	3.78	3.77	*3.59*	**3.87**	3.83	3.82
Soil								
R (pH)	3.64	3.43	*3.36*	3.83	3.84	3.48	3.53	**4.67**
D (aeration)	**4.40**	3.38	3.71	3.54	*3.16*	3.74	3.18	4.00
H (humus content)	*2.00*	**2.92**	2.71	2.74	2.87	2.65	2.82	2.73
F (humidity)	**2.50**	2.40	2.18	2.16	2.29	*2.01*	2.06	2.15
N (nutrient availability)	*1.80*	1.92	1.98	1.93	**2.07**	2.00	2.04	1.92

FIGURES

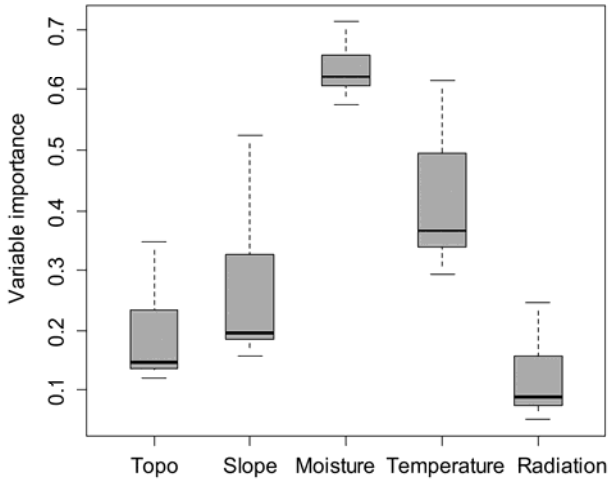


Fig. 1 Importance of each predictor used in the models: a high value (like moisture) indicates an important influence of the predictor in the models. Topo, topographic position; Slope, slope in degrees; Moisture, difference between precipitation and potential evapotranspiration over the growing season (June-August); Temperature, mean temperature for the growing season; Radiation, sum of solar radiation for the growing season.

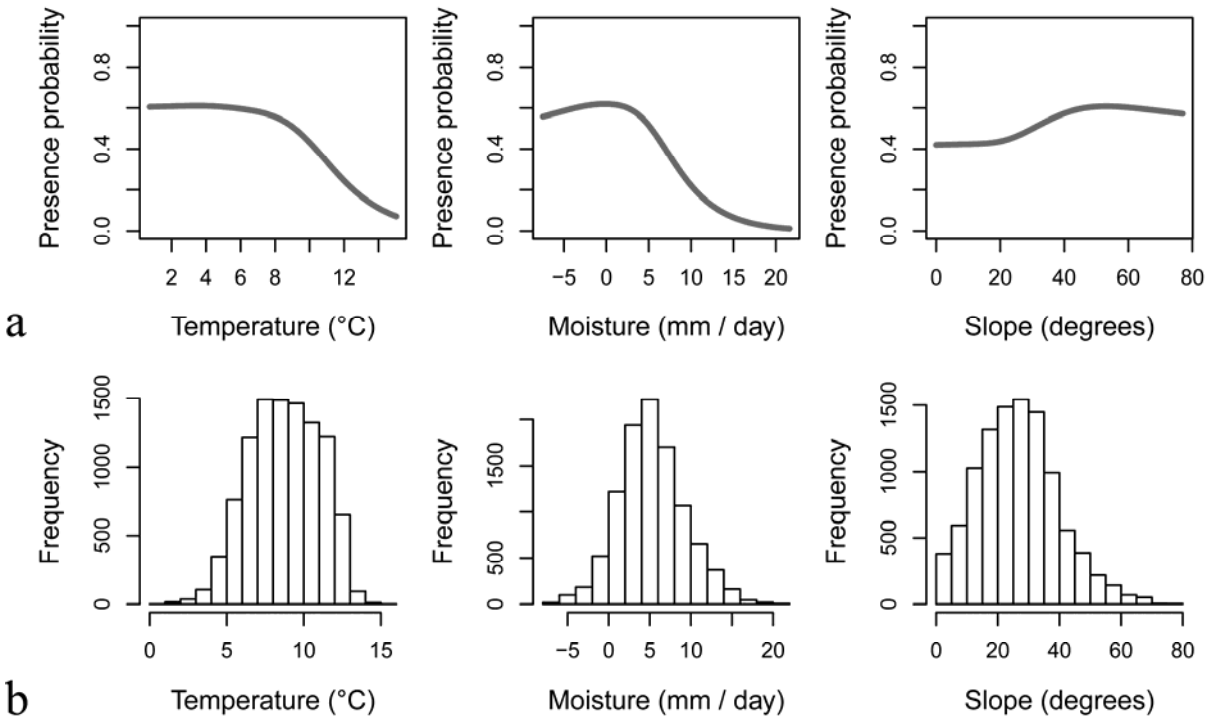


Fig. 2 (a) Response curves for *Leontopodium alpinum* with GAM for the three most important predictors (the response curves for the other models showed similar trends) and (b) frequency distribution of the same predictors in the Swiss Alps. See Fig. 1 for the predictor abbreviations.

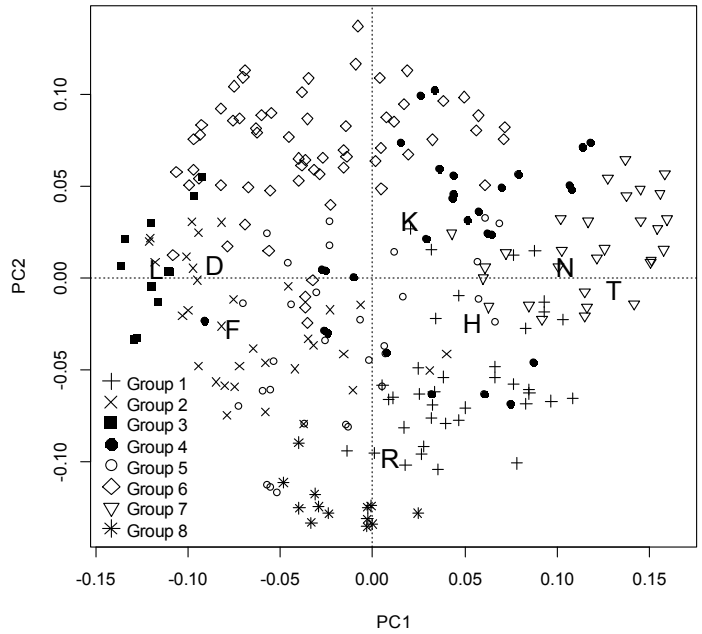


Fig. 3 Principal component analysis with passive projection of the ecological indicator values (Landolt et al. 2010) for climate factors (T, temperature; K, continentality; L, light) and soil characteristics (N, nutrient availability; H, humus content; D, aeration; F, humidity; R, pH). The relevés are represented by symbols related to the groups obtained via clustering analysis (see Table 2). Axes 1 and 2 explain 10.1% and 7.4% of the variance, respectively.

Appendix 1 Predicted habitat suitability map of *Leontopodium alpinum* (red areas) in the Swiss Alps (map from SwissTopo; http://www.swisstopo.admin.ch).

Appendix 2 Synthetic table based on the results of the clustering analysis. Eight groups of relevés were retained. They are approximately ordered following axis 1 of the PCA (see Fig. 3). Species frequency in the groups is given by Roman numeral: V, species frequency >80%; IV, 60-80%; III, 40-60%; II, 20-40%; I, 10-20%; r, <10%. The phytosociological classification of the species is according to Delarze and Gonseth (2008) at the alliance level.

Relevés groups	Phytosociological classification	3	2	6	5	1	4	7	8
Nbre of relevés		13	33	63	30	38	30	29	13
Constant species									
Leontopodium alpinum	Seslerion	V	V	V	V	V	V	V	V
Sesleria caerulea	Seslerion	I	V	IV	V	V	II	V	V
Festuca quadriflora	Caricion firmae	V	V	V	IV	III	III	II	III
Minuartia verna	Seslerion	V	IV	IV	IV	III	II	IV	II
Aster alpinus	Seslerion	I	IV	IV	IV	IV	III	V	II
Agrostis alpina	Elynion	II	V	IV	IV	III	II	IV	II
Elyna myosuroides	Elynion	IV	V	II	III	I	I	I	r
Galium pumilum	Calluno-Genistion	r	r	IV	II	III	r	III	r
Gentiana verna	Seslerion	I	III	IV	II	II	II	III	r
Thymus serpyllum aggr.		r	II	III	III	IV	IV	V	r
Campanula cochleariifolia	Cystopteridion	II	r	I	III	II	r	r	I
Differential species group 3									
Saxifraga oppositifolia	Thlaspion rotundifolii	V	III	II	II	I		r	
Carex rupestris	Caricion firmae	V	I	r	IV	r			
Herniaria alpina	Drabion hoppeanae	IV	r	IV	II		I	II	
Gentiana schleicheri	Drabion hoppeanae	IV	I	III	II				
Linaria alpina s.str.	Thlaspion rotundifolii	III	r	I	r			II	
Draba dubia	Androsacion vandellii	III	r	r	I				
Differential species group 2									
Polygonum viviparum	Caricion curvulae	III	V	r	III	III	r		II
Campanula scheuchzeri			IV	III	II	III	I	I	r
Arenaria ciliata	Elynion	III	III	II	I	r			r
Ligusticum mutellinoides	Elynion	III	III	II		r	r		
Pedicularis verticillata	Seslerion		III	r	I	I			r
Silene acaulis	Caricion firmae		III	r	r	r			I
Differential species group 6									
Draba aizoides	Drabo-Seslerion	IV	II	V	II	I	II	II	
Carex curvula s.l.	Elynion / Caricion curvulae		II	IV	r		I		
Sempervivum arachnoideum	Sedo-Scleranthion	r	II	IV	II	I	III	III	
Artemisia glacialis	Androsacion vandellii			III		r	III		
Veronica fruticans	Festucion variae		r	III	r	r	I	II	
Oxytropis helvetica	Seslerion	II		III	r				
Differential species group 1									
Senecio doronicum	Seslerion		r	III	III	V	II	IV	
Carex sempervirens	Seslerion		III	r	III	V	I	I	r
Anthyllis vulneraria subsp. alpestris	Seslerion		II	I	II	IV	I	II	II
Festuca violacea aggr.	Caricion ferrugineae		r	II	I	III	r	I	
Phyteuma orbiculare	Seslerion		r			III			
Scabiosa lucida	Seslerion		r			III			
Differential species group 7									
Bupleurum ranunculoides s.str.	Seslerion		r	r	II	I	III	V	
Euphorbia cyparissias	Mesobromion		r	II	r	II	II	V	

Carex humilis	Stipo-Poion		r		I	II	r	V	
Oxytropis campestris s.str.	Elynion		II	II	I	II	III	IV	
Dianthus sylvestris	Drabo-Seslerion			r	r	r	III	IV	
Acinos alpinus	Drabo-Seslerion		r	I	r	II	II	IV	
Hieracium pilosella	Mesobromion			I	I	r	II	IV	
Koeleria macrantha	Stipo-Poion			r			II	IV	
Carlina acaulis subsp. caulescens	Mesobromion		r	r	I	II	I	IV	
Teucrium montanum	Stipion calamagrostis			r	I	r	I	IV	r
Astragalus australis	Seslerion			r	I	I	II	III	
Plantago serpentina	Sedo-Scleranthion			r	r	r	II	III	
Pulsatilla halleri	Stipo-Poion			I	r	r	II	III	
Erysimum rhaeticum	Stipo-Poion			r		I	II	III	
Campanula rotundifolia	Asplenion serpentini		r		I	I	I	III	
Trifolium montanum	Cirsio-Brachypodion		r		r	r	r	III	
Astragalus leontinus	Seslerion			r		r	I	III	
Briza media	Mesobromion				r	I	r	III	
Carex caryophyllea	Mesobromion		r			r	r	III	
Galium lucidum	Xerobromion				r	r	I	III	
Dactylis glomerata	Arrhenatherion						r	III	
Linum catharticum						I		III	
Differential species group 8									
Dryas octopetala	Caricion firmae		II	r	I	II			V
Saxifraga caesia	Caricion firmae		r		I	r			V
Carex mucronata	Caricion firmae			r	r	r	I	r	IV
Gentiana clusii	Caricion firmae		I		I	II			IV
Carex firma	Caricion firmae				I	r			IV
Crepis kerneri	Caricion firmae				r				III
Other species									
Helianthemum alpestre	Drabo-Seslerion		II	II	IV	III	II	III	V
Carduus defloratus s.l.	Seslerion		r	I	I	IV	r	IV	r
Galium anisophyllon	Seslerion	r	III	IV	III	II	I	IV	
Euphrasia salisburgensis	Drabo-Seslerion		I	r	II	III	r	II	I
Sedum atratum	Drabo-Seslerion	III	I	I	I	II	I		r
Gypsophila repens	Epilobion fleischeri	I		r	III	II	I	III	r
Euphrasia minima	Caricion curvulae	I	III	II	I	I	I	r	
Helianthemum nummularium s.l.			I	II	II	V	II	V	
Festuca ovina aggr.			I	II	II	II	III	IV	
Lotus corniculatus aggr.			r	I	I	III	I	IV	
Cerastium arvense subsp. strictum	Sedo-Scleranthion		r	III	I	I	III	II	
Gentiana campestris s.str.	Calluno-Genistion		II	I	I	III	r	I	
Hieracium villosum	Seslerion		r		r	III	r	II	I
Globularia cordifolia	Drabo-Seslerion		r		II	II	II	III	II
Juniperus communis subsp. nana	Juniperion nanae		r	I	II	II	I	III	
Silene exscapa	Caricion curvulae	III	II	II	r		I		
Euphrasia alpina	Festucion variae		r	II	I		II	III	
Anthyllis vulneraria subsp. valesiaca	Festucion variae			III	II	r	I	III	
Leucanthemum adustum	Erico-Pinion sylvestris				r	III	r	III	
Botrychium lunaria	Nardion	r	II	II	I	II	r	I	
Carex ericetorum	Elynion	II	I	II	II	r	r	r	
Erigeron alpinus	Seslerion	I	r	II	I	I	r	II	
Gentiana nivalis	Elynion	I	II	I	I	I	r	r	
Hippocrepis comosa	Xerobromion		r	r	r	II	I	I	r
Myosotis alpestris	Caricion ferruginae		r	I	r	II	r	r	r

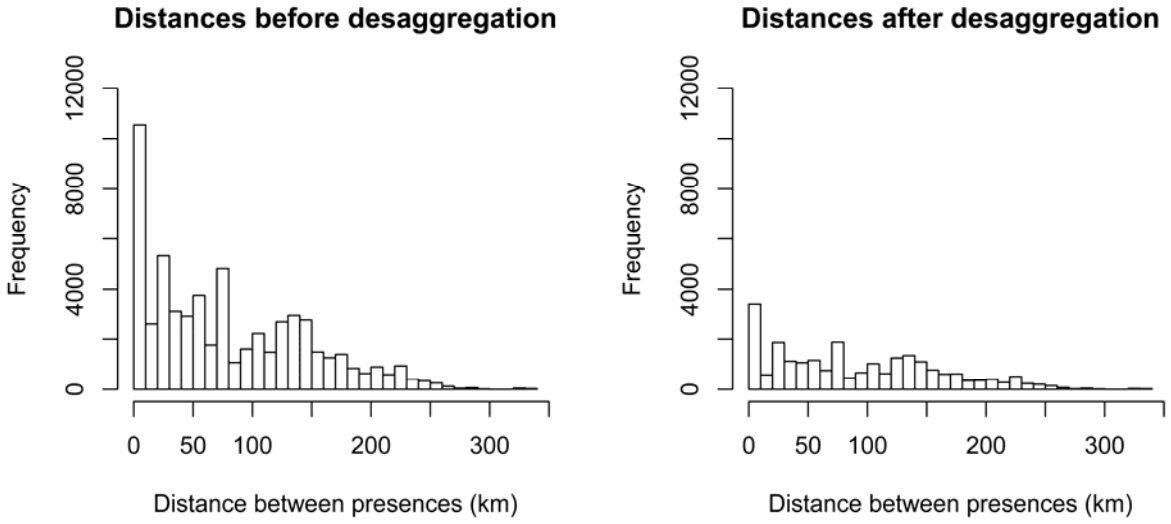
Poa alpina	Poion alpinae	r	II	II	r	II	I		r
Salix serpillifolia	Drabion hoppeanae	II	II	II	I	r	r		I
Saxifraga paniculata	Potentillion		II	II	II	II	II	I	r
Biscutella laevigata	Androsacion vandellii		r		r	II	r	r	II
Calamagrostis varia	Caricion ferruginae		r		r	II	r	r	r
Erigeron uniflorus	Elynion	II	II	r	I	I	r		
Leontodon hispidus s.l.	Poion alpinae		r	r	I	II	II	II	
Potentilla crantzii	Seslerion		II	II	I	I	II	I	
Saxifraga exarata s.l.	Androsacion vandellii	r	r	II	I	r	I		
Trisetum distichophyllum	Petasition paradoxi	r		I	r	r	I	II	
Arctostaphylos uva-ursi	Juniperion sabinae		r	r	I	I	I	r	
Oxytropis halleri s.l.	Elynion		r	r		r	I	r	r
Sempervivum montanum	Sedo-Scleranthion		I	I	I	r	r	r	
Antennaria carpatica	Elynion	I	II	r		r			r
Antennaria dioica	Nardion		I	II	I	r	I		
Anthoxanthum odoratum aggr.	Nardion		r	r	I	r		II	
Chamorchis alpina	Caricion firmae		I	r	r	r			II
Juniperus sabina	Juniperion sabinae			r	r	r	I	II	
Lloydia serotina	Elynion	II	II	r		r			r
Oxytropis lapponica	Elynion		II	r	r	r			r
Pedicularis tuberosa	Festucion variae			II	r	r	r	II	
Potentilla grandiflora	Festucion variae		r		I	r	r	II	
Primula farinosa	Caricion davallianae		II	r	I	I		r	
Pulsatilla vernalis	Caricion curvulae		II	II	r	r	r		
Ranunculus montanus aggr.	Poion alpinae		I	r		II	r	r	
Cirsium acaule	Mesobromion		r	r	r	I		I	
Primula hirsuta	Androsacion vandellii	I	r	r		r	I		
Salix retusa	Arabidion caerulae	r	I	r	r	r			
Arabis ciliata	Seslerion				I	I	r	II	
Artemisia umbelliformis	Androsacion vandellii	II	r	I		r			
Bartsia alpina	Caricion bicolori-atrofuscae		II	r	r	II			
Carex capillaris	Caricion bicolori-atrofuscae	r	II	r			r		
Festuca halleri aggr.	Caricion curvulae	I	I	I			II		
Gentiana brachyphylla	Elynion	r	II	I			r		
Gentiana tenella	Elynion		II	I	r		r		
Helictotrichon versicolor	Caricion curvulae		II	r	r	r			
Minuartia sedoides	Caricion curvulae	II	r	r			I		
Nigritella rhellicani	Nardion		I		r	II		r	
Poa perconcinna	Stipo-Poion			I		r	II	r	
Polygala alpestris	Seslerion			r	r	II		r	
Sempervivum tectorum subsp. alpinum	Drabo-Seslerion		r			II	r	I	
Silene nutans s.str.	Festucion variae			r		I	I	II	
Thesium alpinum	Seslerion		I	r	r	II			
Viola rupestris	Ononido-Pinion		r	II	I		I		
Agrostis rupestris	Caricion curvulae		r	r		r	I		
Allium lusitanicum	Alysso-Sedion				r	r	r	I	
Aster bellidiastrum	Seslerion		r		r	I			r
Carex ornithopoda	Seslerion		I	r		I		r	
Draba siliquosa	Elynion	r	r	I			I		
Festuca varia aggr.	Festucion variae			r	I		I	I	
Kernera saxatilis	Potentillion				I	r		I	r

Luzula spicata s.l.	Caricion curvulae		I	I	I		r		
Parnassia palustris	Caricion davallianae		I		I	I		r	
Pedicularis kerneri	Caricion curvulae		I	r		r	r		
Primula auricula	Potentillion		r			I	r		I
Hieracium piliferum aggr.	Caricion curvulae		r	r		r	r		
Juncus trifidus	Caricion curvulae		r	r		r	r		
Viola pinnata	Erico-Pinion mugo				r	r	r	r	
Androsace chamaejasme	Seslerion		I			I			II
Athamanta cretensis	Petasition paradoxi					r	I		II
Daphne striata	Erico-Pinion mugo				I	II			II
Erica carnea	Ericion				r	I			II
Gentiana utriculosa	Caricion davallianae				r		r	II	
Juncus jacquinii	Caricion curvulae		II	r	r				
Phleum phleoides				r			r	II	
Poa violacea	Festucion variae			r			I	II	
Soldanella alpina	Arabidion caerulae		II	r		I			
Taraxacum laevigatum aggr.				I	I		II		
Thalictrum foetidum	Geranion sanguinei				r		r	II	
Trifolium pratense s.l.	Arrhenatherion				r	I		II	
Achillea millefolium aggr.	Arrhenatherion					r	r	I	
Draba fladnizensis	Drabion hoppeanae	I	r	r					
Festuca arundinacea s.str.	Agropyro-Rumicion			r	r			I	
Hieracium angustifolium	Caricion curvulae			I	I	r			
Homogyne alpina	Vaccinio-Piceion		I	r		I			
Minuartia mutabilis				r			I	I	
Minuartia recurva	Caricion curvulae			r		r	I		
Myosotis stricta	Sedo-Veronicion			r	r			I	
Phyteuma hemisphaericum	Caricion curvulae		I	r		r			
Rhamnus pumila	Potentillion				r	r	I		
Saxifraga bryoides	Androsacion alpinae		I			r	r		
Selaginella selaginoides	Caricion bicolori-atrofuscae		I	r		r			
Senecio incanus s.str.	Caricion curvulae			I		r	r		
Silene rupestris	Sedo- Scleranthion			r	I			r	
Silene vulgaris s.str.	Arrhenatherion				r	I		I	
Stipa eriocaulis s.str.	Stipo-Poion		r				I	I	
Thalictrum minus s.l.	Geranion sanguinei		r			r	I		
Artemisia genipi		r	r	r					
Carex liparocarpos	Stipo-Poion			r			r	r	
Carex ornithopodioides	Arabidion caerulae		r		r	r			
Centaurea scabiosa s.l.	Mesobromion					r	r	r	
Cirsium spinosissimum	Rumicion alpinii				r	r	r		
Gentiana engadinensis	Seslerion caeruleae		r		r	r			
Luzula lutea	Caricion curvulae		r	r	r				
Oxytropis fetida	Drabion hoppeanae	r		r			r		
Phyteuma globulariifolium s.l.	Caricion curvulae	r		r			r		
Poa bulbosa	Sedo-Veronicion			r			r	r	
Polygala alpina	Caricion firmae			r		r		r	
Potentilla aurea	Nardion		r	r		r			
Silene vulgaris subsp. glareosa	Petasition paradoxi			r		r	r		
Solidago virgaurea subsp. minuta	Epilobion angustifolii			r	r	r			
Taraxacum dissectum	Caricion firmae			r			r	r	
Veronica fruticulosa						r	r	r	
Astragalus sempervirens	Drabo-Seslerion			r			II		

Hedysarum hedysaroides	Caricion ferruginae		r			II						
Hieracium bifidum aggr.	Seslerion					II						II
Scabiosa columbaria s.str.	Mesobromion							r		II		
Anthyllis vulneraria s.str.	Mesobromion				r							I
Coeloglossum viride	Nardion		r			I						
Erigeron neglectus	Elynion			r		I						
Festuca rubra aggr.	Cynosurion		r			I						
Helictotrichon pubescens	Arrhenatherion					r				I		
Hieracium tomentosum	Potentillion							I		r		
Koeleria vallesiana	Stipo-Poion							I		r		
Plantago alpina	Nardion					r		I				
Poa laxa	Androsacion alpinae	r						I				
Polygala chamaebuxus	Ericion					I		r				
Potentilla frigida	Caricion curvulae		I	r								
Pulsatilla alpina s.str.	Caricion ferruginae		r			I						
Rhinanthus minor	Molinion					r				I		
Saussurea alpina s.str.	Elynion		I			r						
Stachys recta s.str.	Xerobromion					r				I		
Stipa pennata	Cirsio-Brachypodion							r		I		
Androsace obtusifolia	Caricion curvulae		r	r								
Androsace vitaliana	Androsacion alpinae			r				r				
Astragalus frigidus	Caricion ferruginae		r			r						
Carex nigra	Caricion fuscae		r	r								
Cotoneaster integerrimus	Juniperion nanae					r		r				
Dianthus glacialis	Elynion		r									r
Echium vulgare	Dauco-Melilotion							r		r		
Erigeron glabratus	Seslerion		r	r								
Gentiana orbicularis	Drabion hoppeanae		r			r						
Gentiana ramosa	Festucion variae			r	r							
Geum montanum	Nardion			r		r						
Laserpitium siler	Stipion calamagrostis					r		r				
Leontodon helveticus	Nardion		r			r						
Leontodon incanus s.str.	Erico-Pinion sylvestris					r		r				
Linum alpinum	Seslerion		r			r						
Oxytropis jacquinii	Seslerion		r			r						
Pinus cembra			r			r						
Plantago atrata s.str.	Poion alpinae		r			r						
Poa nemoralis	Tilion platyphylli							r		r		
Salix reticulata	Arabidion caerulae		r			r						
Taraxacum alpinum aggr.	Arabidion caerulae			r	r							
Tephroseris capitata	Seslerion		r			r						
Veronica aphylla			r			r						
Viola calcarata	Poion alpinae					r						r
Androsace puberula	Caricion curvulae			II								
Astragalus monspessulanus	Ononido-Pinion									II		
Cuscuta epithymum	Caucalidion									II		
Pinguicula alpina	Cratoneurion											II
Scutellaria alpina	Seslerion							II				
Achillea erba-rotta subsp. moschata	Androsacion alpinae							I				
Achillea nana	Drabion hoppeanae			I								
Asplenium ruta-muraria	Potentillion							I				
Campanula thyrsoides	Caricion ferruginae					I						
Carex atrata s.str.	Elynion		I									

Carex fimbriata	Caricion curvulae						I		
Carex flacca	Molinion					I			
Crepis alpestris	Erico-Pinion mugo					I			
Daphne mezereum	Cephalanthero-Fagenion					I			
Dianthus caryophyllus	Festucion variae					I			
Gymnadenia odoratissima	Molinio-Pinion					I			
Helictotrichon parlatorei							I		
Hieracium hoppeanum	Festucion variae					I			
Myosotis arvensis	Caucalidion							I	
Onobrychis montana	Seslerion						I		
Phleum hirsutum	Caricion ferruginae					I			
Poa molinerii	Alysso-Sedion						I		
Potentilla caulescens	Potentillion						I		
Rhinanthus glacialis	Nardion					I			
Sanguisorba minor s.str.	Mesobromion							I	
Valeriana montana	Petasition paradoxi					I			
Alchemilla conjuncta aggr.	Seslerion					r			
Anemone baldensis	Thlaspion rotundifolii			r					
Artemisia borealis	Drabion hoppeanae			r					
Luzula multiflora						r			
Potentilla erecta	Nardion					r			
Stemmacantha rhapontica s.str.	Calamagrostion					r			
Mean number of species		16.6	29	29.3	24.7	35.1	24.6	40.8	14.6

Appendix 3 Distance between pairs of occurrences in the original data file and after desaggregation to avoid autocorrelation with minimal distances of 250 m between occurrences.



REMERCIEMENTS

Je voudrais remercier ici les nombreuses personnes qui ont joué un rôle dans l'accomplissement de cette thèse.

Tout d'abord, mon directeur de thèse, le professeur *Antoine Guisan* pour m'avoir permis de réaliser ce travail, pour ses conseils avisés, ses idées et la liberté très appréciable dont il m'a laissé profité.

Les membres de mon comité de thèse, le professeur *Stefan Kunz*, président; les experts, *Miska Luoto* et *Jean-Claude Gégout*, qui ont accepté d'évaluer ce travail et m'ont apporté de nouveaux points de vues intéressants, ainsi que *Wilfried Thuiller* et *Martin Zobel* qui m'avaient donné de bon conseils lors de mon comité de demi-thèse.

Pascal Vittoz, également expert de ce travail, à qui je dois mes connaissances botaniques, qui a énormément aidé à la collecte des données de terrain et qui m'a donné de nombreux conseils tout au long de cette thèse.

Les membres actuels et anciens du groupe Ecospat:

Julien Pottier qui m'a transmis un précieux savoir en écologie des communautés ainsi qu'en code R. Merci aussi d'avoir instauré le café de 10h, on le respecte toujours.

Christophe Randin, qui m'a fait découvrir le travail de terrain et la magnifique région des Préalpes vaudoises et avec qui j'ai eu de passionnantes discussions.

Loïc Pellissier, mon collègue de projet, pour toutes les discussions, le temps passé sur le terrain et à relire mes articles. La collaboration fut vraiment agréable. Merci également de m'avoir amené à la foire du Valais.

Jean-Nicolas Pradervand avec qui j'ai eu le plaisir de passer beaucoup de temps sur le terrain ainsi qu'au bureau pour le dernier chapitre de cette thèse.

Olivier Broennimann, Carmen Cianfrani, Robin Engler, Luigi Maiorano, *Charlotte Ndiribe, Blaise Petitpierre*, *Eric Pinto, Dorothea Pio*, *Patricio Pliscoff*. Grâce à vous c'était toujours un plaisir d'être au travail. Merci pour les discussions, plus ou moins sérieuses, et le bon temps qu'on a pu passer ensemble durant les « ecospartys » et ailleurs.

Vanessa Rion, *Sara Giovanettina*, *Leila Rossier*, *Mélanie Ischer* et *Coralie Boheler* dont j'ai eu l'opportunité d'encadrer le travail de master. Vous m'avez apporté une aide indispensable et un regard critique aussi bien pour la récolte de données que pour les analyses statistiques.

J'ai eu la chance de participer à trois étés de récolte de données avec *Saskia Godat, Christian Purro, Virginie Favre, Aliocha Coiana, Valéry Uldry, Alexander von Ungern-Sternberg, Philippe Chatelain, Aurore Gelin, Alain Reymond* et *Vincent Sonnay*. Merci pour l'énorme travail que vous avez accompli, les échanges de connaissances, la slackline et la pétanque dans le jardin, les soupers et les discussions du soir autour d'un verre de prune.

Séverine Wider, *Tania de Watteville* et *Maurice de Watteville* qui en mettant à notre disposition leur chalet de Villars nous ont permis de profiter de saisons de terrain cinq étoiles.

François Bonnet qui a accueilli notre «labo de terrain » au chalet du jardin botanique de Pont de Nant durant trois étés.

Nicole Galland qui m'a fait découvrir et aimer la botanique durant ma première année d'étude, ainsi que les membres du groupe botanique de la *Murithienne* avec qui j'ai pu développer cette passion.

Félicidad Jacquiéry, *Giusy Métraux, Nadia Bruyndonckx* et *Franck Chalard* pour toute l'aide administrative et technique.

Nadine, *Emmanuelle*, *Natacha, Fanny* pour m'avoir permis de parler d'autre chose que de science, échanger des livres, courir, boire du thé ou des mojitos et manger des pâtes au pesto.

Elisabeth et *Steve Litsios* pour leur accueil toujours chaleureux à la Chaux de Fonds.

Toute ma famille, et plus particulièrement ma sœur *Hélène*, mon frère *Pierre-Louis*, mes parents *Anne-Dominique* et *Bernard* pour leur soutien, leurs encouragements et m'avoir permis de garder les pieds sur terre.

Finalement, merci *Glenn*, pour ta présence, ton soutien et tout le reste, je me réjouis de la suite !