UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole de Biologie

MODELLING THE PLANT SPECIES RICHNESS: A COMPARISON BETWEEN TWO APPROACHES

Travail de Maîtrise universitaire ès Sciences en comportement, évolution et conservation

Master Thesis of Science in Behaviour, Evolution and Conservation

par

Vanessa RION

Directeur : Prof. Antoine Guisan

Superviseur (s) : Prof. Antoine Guisan, Anne Dubuis

Expert : Prof. Martin Zobel

Département d'écologie et d'évolution (DEE)

Février, 2010

Abstract

Plant biodiversity is predicted to be negatively influenced by climate change and human activities. Plant species richness modelling is an important tool in conservation biology allowing predicting and following the biodiversity response to changing environment. Highlighting biodiversity hotspots, it allows elaborating conservation strategies to conserve the current species richness.

Two modelling approaches are currently used in this research field. In the first approach called macroecological or direct approach, the species number is directly predicted from observed species counts through a set of topo-climatic predictors. The second approach, called cumulative approach consists in pilling up independent species distribution models (SDMs): each species presences and absences are linked to the environmental predictors, giving a potential distribution map. These potential distributions are then pilled up to generate a species richness map.

In this study, I create a plant species richness map for the Western Swiss Alps with both approaches and compare the obtained ouputs.

The results reveal that the cumulative approach clearly tends to overpredict the species number at every altitude. This trend is more important at low elevation. The SDMs used in this approach fit a larger niche than the true realized one because not all factors limiting the distribution of each species are taken into account. These errors accumulate during the pilling of the potential distribution maps, inducing the overprediction of the total species number in each pixel. In contrast, the direct approach is more restrictive and predicts a smaller species number at these low altitudes. It also illustrates that plant species richness at lower altitudes seems to undergo a negative human influence.

Besides, both approaches show the same trends: less species at very high altitude (>2500 m a.s.l) and the highest species richness at intermediate altitudes (at around 1500 m a.s.l.). These modelling techniques have both strengths and weaknesses, but seem to be complementary. Thus, they should be used together in biodiversity modelling.

Résumé

La diversité des plantes est influencée négativement par le changement climatique et les activités humaines. Les modèles de richesse spécifique végétale sont donc des outils importants en biologie de la conservation, car ils permettent de prédire et de suivre l'évolution de la biodiversité face à un environnement en mutation. Ils permettent également de détecter les endroits à haute biodiversité (hotspots) et donc de définir des priorités pour l'application des stratégies de conservation visant à maintenir la richesse spécifique actuelle.
Deux approches de modélisation sont couramment utilisées dans ce domaine. Dans la première, appelée approche macroécologique ou approche directe, le nombre total d'espèces est directement prédit à partir du nombre d'espèces observées sur le terrain et d'un jeu de prédicteurs topographiques et climatiques. La seconde, appelée approche cumulative, consiste à empiler des distributions potentielles d'espèces : les présences et absences de chaque espèce sont liées aux prédicteurs environnementaux, donnant une carte de distribution potentielle pour chaque espèce. Ces distributions sont ensuite empilées et additionnées afin de générer une carte de richesse spécifique.

Dans cette étude, j'ai comparé ces deux techniques et plus précisément les résultats obtenus en appliquant ces deux approches à la richesse spécifique des plantes des Préalpes vaudoises, dans les Alpes occidentales suisses.

L'approche cumulative tend clairement à surprédire le nombre d'espèces. Cette surpédiction semble être plus importante dans les régions de basse altitude. Les modèles de distribution d'espèce individuels prédisent une niche plus large que la niche réelle car tous les facteurs limitant la distribution de l'espèce ne sont pas pris en compte dans le modèle. Cette erreur s'accumule lors de l'addition des modèles, ce qui entraîne la surprédiction du nombre d'espèces. Au contraire, l'approche directe est plus restrictive et prédit une richesse spécifique végétale plus basse aux faibles altitudes. Ceci semble mettre en évidence l'influence négative de l'homme sur la biodiversité végétale.
Ces deux approches montrent cependant les mêmes tendances: moins d'espèces aux très hautes altitudes (>2500 m.) et un maximum d'espèces aux altitudes intermédiaires soit aux alentours de 1400 mètres d'altitude.
Ces deux techniques permettant de modéliser la richesse spécifique, possèdent donc chacune des avantages et des inconvénients différents, mais semblent être complémentaires. Le futur de la modélisation de la biodiversité pourrait résider dans l'utilisation conjointe de ces deux techniques.

Introduction

Plant biodiversity can be defined as the diversity of a plant assemblage within a given unit or area. It can be the simple count of species – species richness (Margules, 1989) – or more complex measures accounting for individual species abundances, species traits or any genetic information such as relatedness between species. Species richness can reflect on an ecosystem's vitality (Tilman, 1996; Folke et al., 1996), for instance with plant species disappearance possibly inducing the loss of animal species depending on them, resulting in a threat to the whole ecosystem stability. Such risk is usually considered to be minimized when the diversity is high, because then several species can ensure the same functional role (some species are interchangeable) and if one disappears, another takes on to maintain a vital ecosystem function (insurance hypothesis; Yachi & Loreau, 1999). It is thus desirable and human responsibility to maintain as much biological diversity as possible in any ecosystem. However, human activities and associated global environmental changes now represent a growing threat to biodiversity and recent forecasts anticipate dramatic species loss in response to future climate changes (e.g. Thuiller et al., 2005 for European plants).

Mountain areas worldwide are important hotspots of biodiversity and endemism. Yet, mountain ecosystems are also considered especially vulnerable and exposed to climate change (Guisan et al., 1995; Theurillat et al., 1998; Diaz et al., 2003; Nogués-Bravo et al., 2007). Thus biodiversity loss is expected in mountain areas in response to climate change (Thuiller et al., 2005; Randin et al., 2009a; Engler et al., in prep.). In addition to the forecasted temperature increases (from +1.1°C to +6.4° depending on the scenario according to the Intergovernmental Panel on Climate Change), climate change will also modify the water regime (Diaz et al., 2003). In the Swiss Alps, the winter precipitations should decrease at lower altitude and increase at very high altitudes, whereas rainfalls in summer will tend to decrease at every elevation (Beniston, 2006). The quantity of snow as well as its duration will be reduced, and glaciers will melt more rapidly (Diaz et al., 2003; Beniston, 2006). These changes will increase the potential extinction risk of many indigenous plants species in these areas (Guisan & Theurillat, 2000b; Dirnböck et al., 2003), affecting negatively plants communities (Guisan et al., 1995; Guisan & Theurillat, 2000b; Walther, 2003) and plant biodiversity patterns (Saetersdal et al., 1998; Thuiller et al., 2005).

In the Alps, global warming has already induced some observable changes on plants: treeline is slightly but perceptibly shifting upper in altitude (Gehrig-Fasel et al., 2007; Vittoz et al., 2008), species migrations toward higher altitudes or northern latitudes have been observed (*Viscum album*, Dobbertin et al., 2005; *Ilex aquifolium*, Walther et al., 2005b; Vittoz et al., 2006), and exotic species have become established at lower altitude and tend to replace the indigenous flora (Walther, 2002). Plant species richness in the subalpine belt and in the top of the mountains is predicted to increase (Pauli et al., 1996; Walther et al., 2005a;

Vittoz et al., 2006; Pauli et al., 2007). Guisan & Theurillat (2000b), Engler et al. (2009) and Randin et al. (2009a) found that favourable habitats of many alpine and subalpine species are expected to decrease in their extent, inducing a loss of alpine plant diversity. The risk of extinction will be greater for strict alpine and nival species than for lower elevation species or species that tolerate a larger range of altitudes (Guisan & Theurillat, 2000b).

Species richness modelling is an important tool in conservation biology. It enables to know were biodiversity hotspots are and consequently to elaborate conservation strategies to preserve the current species richness threatened by climate change and anthropogenic activities (Austin, 1999). Consequently it is important to model the plant biodiversity in order to anticipate how it will change in the future. Species richness can be modelled in two different ways (Ferrier & Guisan, 2006). The first approach, called “assemble first, predict later” (direct approach in this study), corresponds to modelling species richness per se (i.e. as a final community property). The second approach, called “predict first, assemble later” (cumulative approach in this study), corresponds to modelling individual species first, and then assembling them to recompose species richness.

In the first case, the number of species in inventoried plots is directly related through a statistical model to a set of environmental predictors available for the given area, that are considered as important factors for plant development (Pearson et al., 2002), resulting in a direct prediction of species richness patterns (e.g. Thuiller et al., in press). These factors are also believed to be the main factors determining the species richness distribution in mountain areas (Pausas & Austin, 2001; Marini et al., 2008). This direct approach is also called macroecological approach (Guisan & Rahbek, in prep.) and is based on macroecological hypotheses such as described in Currie (1991) or Whittaker et al. (2001). More precisely, species richness patterns are determined by some important environmental and biotic variables, as for example the amount of available nutrients, water and energy (light), temperatures, environmental heterogeneity and the level of disturbance (Pausas & Austin, 2001). For instance, a first hypothesis is that the amount of available energy limits species richness. However, it is often expected that the highest number of species is in regions characterized by the greatest amount of available energy (Currie, 1991; Pausas & Austin, 2001). A second hypothesis is that a very heterogeneous habitat shelters more species than a homogeneous one because it provides more ecological niches (Pausas & Austin, 2001; Dufour, 2006). A third hypothesis states that extreme and varying conditions, as for example the lack of resources, limit the number of species, because few species are physiologically adapted to stressful and changing conditions (Currie, 1991; Whittaker et al., 2001). A fourth hypothesis states that a too small or too high disturbance rate reduces the number of species (Pausas & Austin, 2001). A small disturbance level selects more competitive species, ultimately only allowing the most dominant one to survive. On the other end of the gradient, a

high disturbance rate prevents late- and mid-successional species to replace early-successional ones (i.e. pioneers). The maximum of species richness is thus found at intermediate level of disturbance, where species from many successional stages can co-exist (Connell, 1978). Concerning the amount of nutrients, species richness is expected to be higher in regions that have intermediate level of nutrients (Pausas & Austin, 2001). Indeed, a too small amount of nutrients limits the species number because not very much species are adapted to this extreme condition. Furthermore, a too high level of nutrients decreases the species richness too, because in this condition, only the most competitive species become dominant, supplanting all the others plants (Pausas & Austin, 2001). Concerning the temperatures and the water availability, species richness is expected to increase with these two factors (Pausas & Austin, 2001). Generally, altitude is used as substitute of the temperature variable, because it reflects well this factor. Thus, species richness is predicted to decrease with altitude (Austrheim, 2002; Theurillat et al., 2003). However this tendency may not be valid in all situations, because it can be affected by other variables (Pausas & Austin, 2001). See Guisan & Rahbek (in prep) or Pausas & Austin (2001) for an extensive summary.

The second approach to model species richness, called cumulative approach in this study, is more recent and based on individual species distribution models (SDMs; Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005). Presences and absences of each species are statistically related to a set of environmental predictors for the area that are also believed to be essential variables determining plant growth and reproduction, and thus related to energy and nutrients (Pearson et al., 2002). The fitted relationship is the model used then to predict the potential distribution of the species' suitable habitats, with the envelope of all suitable habitats representing in fine the species' realized environmental niche (sensu Hutchinson, 1957). All potential maps are then pilled up (cumulated) in order to produce a species richness map (e.g. Parviainen et al., 2009).

Hence, both approaches generate comparable prediction. So far, they were used separately in different studies in order to model animal or plant species richness (Lobo & Martín-Piera, 2002; Gutiérrez Teira & Peco, 2003; Lobo et al., 2004; Thuiller et al., 2006; Bergamini et al., 2007; Kissling et al., 2007; Rahbek et al., 2007; Gotelli et al., 2009 for the macroecological approach; Lehmann et al., 2002; Olden, 2003; Peppler-Lisbach & Schröder, 2004; Guisan & Thuiller, 2005; Schmidt et al., 2008; Parviainen et al., 2009 for the cumulative approach). Few studies attempted to compare them using a same dataset (e.g. Guisan & Theurillat, 2000b; Algar et al., 2009; Steinmann et al., 2009).

Here, we used a large comprehensive dataset of vegetation plots sampled across a rather well studied area in the Swiss Western Alps (Drake et al., 2006; Engler et al., 2009;

Randin et al., 2009a; Randin et al., 2009b) to implement and compare both approaches under present climatic conditions, discussing their strengths and weaknesses.

Material and methods

Study area

The study area is located in the Western Swiss Alps (46°10' to 46°30'N; 6°60' to 7°10'E). It covers an area of ca. 700 km^2 (Figure 1). Elevation ranges from ca. 400m a.s.l. in the bottom of the Rhône Valley to 3210 m a.s.l. on the top of the Diablerets massif. Another major mountain in the area is the Muveran massif (3051 m a.s.l.). Some important glaciers are present in this zone. The climate is temperate: rainfalls increase with elevation (the annual precipitation fluctuate from 1200 mm to 2600 mm between 600 m a.s.l. and 3000 m a.s.l.; Bouët, 1985), while temperatures decrease (the annual mean temperatures vary from ca. 8°C at 600m a.s.l. to -5°C at 3000m a.s.l.; Bouët, 1985). The vegetation sequence along elevation is the following: the colline belt (until 800 m a.s.l.) is composed of deciduous forests (represented principally by *Fagus sylvatica* L.), the montane belt (800 to 1600 m a.s.l.) consists of mixed forests (*F. sylvatica* and *Abies alba* Mill.), the subalpine belt (1600 to 2400 m a.s.l.) of coniferous forests (*Picea abies* Karsten), the alpine belt (2400 to 3000 m a.s.l.) of grassland, meadow and heath vegetation and the nival belt (>3000 m a.s.l.) is composed of small and sparse vegetation. The area contains some important natural reserves, such as the one in Vallon de Nant (commune of Bex). Human also influences vegetation in the study area: many non-forested areas are used as pastures or mowed, and fertilization is important from the lowlands to the subalpine belt.

Vegetation data

The vegetation data originate from different field sampling campaigns conducted in the study area. In total, 912 points were visited. The last field sampling was done in the summer 2009, during which 299 new vegetations plot were sampled to complement existing vegetation data from previous field samplings. All plots were chosen according to a random-stratified sampling design (Maggini et al., 2002). This strategy was chosen because, in a simulation study, it proves more efficient than other strategies for predicting species distribution (Hirzel & Guisan, 2002). The stratification was done by elevation, slope and aspect, and the sampling was limited to the open and non-woody vegetation only, i.e. grasslands, meadows, rocks and screes. Furthermore, each selected point was separated from others by a minimum distance of 200 m to avoid spatial autocorrelation.

In the field, the exact coordinates of the vegetation plots were found with a Trimble GPS receiver, with an accuracy of 1 m. At each visited location, a grid of 2m x 2m was placed on the ground, the bottom left of the grid corresponding to the target coordinate. All

plant species in this 4 m^2 square were recorded, as well as their respective abundance according to the Braun-Blanquet (1964) modified scale (Table 1). Percents of vegetation density, bared ground, rocks and mosses were estimated in the whole plot. Then, in each plot, five smaller subplots (20 cm x 20 cm) were sampled, one close to each plot angle and one in the centre of the plot (Figure 2): all the plants present in these subplots were inventoried as well as their respective cover, using a scale from 0 to 100%. Furthermore, in each of these subplots, vegetation height was measured at five places to obtain a good estimate for the whole plot. If the plot position was not accessible or was covered with shrubs, it was displaced (using a standard random procedure). Moreover, samples of soil were taken in each subplot for further analyses of physical characteristics and DNA content.

Only the species that had more than 20 occurrences throughout all sampled plots were kept for further analyses, as below this threshold it becomes difficult to fit models. So, in total, 260 plant species were used for the modelling (Appendix 1).

Climatic and topographic predictors

In this study, three climatic predictors (degree-days, moisture index and global solar radiations) and two topographic predictors (slope and topographic position) were used for fitting models (Table 2; Appendix 2). Climatic predictors corresponded to monthly values of the period 1961-1990. Methods used to make these predictors are described in Zimmermann & Kienast (1999) and Zimmermann et al. (2009).

These predictors are linked to energy and nutrients and are thought to be the most meaningful variables for plant growth and reproduction (Pearson et al., 2002) and represent some important factors determining the species richness pattern (Pausas & Austin, 2001). They have already been used and tested in several other modelling studies in mountains areas (Randin et al., 2006; Engler et al., 2009; Randin et al., 2009b).

Degree-days represent the energy necessary to a plant to accomplish its life cycle. In this study, a threshold value of 0°C (minimum temperature for growth) was used because it is typically the minimum temperature for plant growth at high altitude (Lenihan, 1993). The moisture index, calculated as the difference between precipitations and potential evapotranspiration, represents the soil water availability. This index has an important influence on the growth of plants, because it expresses the amount of water present in the soil that is potentially available for plants. In this study, only the monthly averages from June to August were used, because during these three warm and dry months, moisture is a limiting factor to plant growth. Global solar radiations were calculated for the whole year. This predictor has an energetic unit (kJ) and is an important variable for plants since photosynthesis depends on the amount of light available. The two topographic variables, slope and topographic position, were derived from the Swiss digital elevation model (DEM) at a resolution of 25 m, thus representing the relief of the study area. Slope is both a direct

factor through gravimetric effects on plants and an indirect measure of water accumulation and of the incidence angle of solar radiations. Topographic position represents the concavity or convexity of mountain slopes. It is derived from a mobile window of increasing size applied to the DEM. Positive values of topographic position indicate tops of mountains and ridges, while negative values correspond to valleys, plains or toe slopes. All these predictors were available at the 25 m resolution.

Correlations between these variables were tested by principal components analysis (PCA) and correlation matrix (Appendix 3). To enter two variables in a same model, their pairwise correlation had to be smaller than 0.7.

Modelling approaches to species richness

Model calibration and all other statistical analyses were computed in version 2.9.0 of the R software (R Development Core Team, 2009). These models were then projected in ArcGIS (ESRI, 2008) in order to produce species richness maps.

A scheme of the approaches used in this study is represented in Figure 3. A map of the observed species richness was also created with the number of species inventoried in each of the 912 plots.

Direct modelling of species richness

In this direct approach, species richness itself is the response variable in the models. Species richness at each of the 912 sites can be simply calculated by summing all the species inventoried in each plot. In this study, however, only the 260 most frequent species were considered, to match the analyses in the next section (cumulative approach). As a count of species, it is best modelled as Poisson or negative-binomial probability distributions (Fisher et al., 1943; Vincent & Haworth, 1983; Guisan & Zimmermann, 2000). The zero-inflated negative binomial distribution is a part of the negative-binomial probability distribution that was developed and adapted to data with too much zero counts numbers (Welsh et al., 1996). Two statistical approaches then allow considering these probability distributions: generalized linear models (GLMs; McCullagh & Nedler, 1989) and generalized additive models (GAMs; Hastie & Tibshirani, 1990). GLMs are a part of regression models, in which the response variable is related to one (simple regression) or more predictors (multiple regression) through a link function depending on the type of distribution used. In the other hand, GAMs resembles to GLMs but they assume that the response variable is linked to the environmental predictors with a non-linear and non-parametric smoothing function. The Poisson distribution is the most used distribution for counts data (Guisan & Zimmermann, 2000; Guisan & Theurillat, 2000a). The negative binomial distribution is less frequent in ecology and could better apply to our data because a lot of the sampling plots were composed of a very small number of species (Appendix 4). In this study, both distributions

(Poisson and zero-inflated negative binomial) were used. The zero-inflated negative binomial was performed here using the library VGAM in R. Furthermore, for both distributions, GLMs were fitted with linear and quadratic terms for each predictors, and GAMs with four degrees of freedom. A logarithmic link function was used for the Poisson distribution. An Akaike information criterion (AIC; Akaike, 1973) stepwise procedure in both directions was applied to Poisson models in order to select the most important predictors. With this procedure, the five topo-climatic predictors were selected. No stepwise selection was applied to negative-binomial models. Thus, the five predictors selected were inserted in all models. Species richness predictions maps were prepared in ArcGIS 9.3 (ESRI, 2008). The four species richness models (GLM and GAM with Poisson and negative binomial distributions) were projected onto the whole study area in ArcGIS 9.3. A mask containing forests, lakes, urbanized zones, roads, rivers and glaciers was used to filter each of the projection in order to avoid making predictions in fully unsuitable habitat.

Predictions from Poisson and negative-binomial models were compared with a spearman rank correlation in order to see if these both distributions give similar results or species richness patterns. Furthermore, for each distribution, a comparison between the species richness predicted by GLM and the GAM was also performed with a spearman rank correlation.

The accuracy of each model was evaluated by a repeated split-sample procedure (a special case of two-fold cross-validation). In this procedure, an evaluation data set was obtained by randomly splitting the original data set into two 70%-30% partitions, the 70% partition serving at fitting the models, while the 30% left-out data was used for independent evaluation. For each split-sample repetition and for each model, a spearman correlation between observed and predicted species richness was performed on the 30% evaluation data set. Measures of deviance and AIC were obtained for each of the 100 runs. Deviance reduction (D^2) – a measure of explained variance in maximum likelihood methods –, the adjusted D^2 – the equivalent to D^2 but with the number of observations (n) and the number of predictors (p) taken into account (penalty; Guisan & Zimmermann, 2000) – and the ΔAIC were calculated for all models by the following formulas:

$$D^2 = \frac{\left(\text{Null deviance} - \text{Residual deviance}\right)}{\text{Null deviance}} \qquad \text{adjusted } D^2 = 1 - \left[\frac{(n-1)}{(n-p)}\right] \cdot \left[1 - D^2\right]$$

$$\Delta\text{AIC} = \text{AIC null model} - \text{AIC model}$$

Cumulative modelling of species richness

In this cumulative modelling approach, presence-absence of each species are first modelled individually and their binary predictions are then pilled-up to reconstruct species richness. Presence-absence models and predictions were built for the 260 most frequent plant species across the study area with the BIOMOD library in R (Thuiller, 2003; Thuiller et al., 2009a; Thuiller et al., 2009b). Six different modelling techniques were used within an ENSEMBLE forecasting approach (Araújo & New, 2007): GAM (Generalized Additive Model), GLM (Generalized Linear Model), GBM (Generalized Boosting Model), RF (RandomForest), CTA (Classification Tree Analysis) and MARS (Multiple Adaptive Regression Splines). GBMs resemble to GLMs but use and combine more models, giving more robust predictions. RF use an algorithm that allow the construction of classification trees. CTA is an alternative to regression models, using trees building. MARS models are non-parametric regression and take all effects of the predictors into account, as additive and interactions effects. For more details about the different techniques, see Marmion et al., 2009. For each model and for all pixels in the study area, the presence or the absence of each species was measured across the 912 plots. The ensemble forecasting approach only included models of sufficient accuracy, i.e. those that have an AUC (area under the receiver-operating characteristics curve) greater or equal to 0.7, as this is the threshold value above which models can be considered as relevant (Swets, 1988). A different weight was then assigned to each retained model, which was proportional to its accuracy as measured by its AUC (linear weighting function). Ensemble predictions were calculated as average probabilities across all models and were transformed into presence-absence (1-0) maps. Finally, all binary ensemble species prediction maps were pilled-up resulting in a species richness map for the whole study area. The accuracy of this cumulative species richness map was evaluated by a spearman correlation between observed and the predicted species richness.

Comparison between the two approaches

The two modelling approaches - direct versus cumulative - were compared quantitatively with a spearman rank correlation and visually by looking at the prediction maps.

Furthermore, the plots of the residuals of both approaches – calculated as observed minus predicted species number – as function of the altitude and the plots of the species number predicted by each approach according to altitude were also used to compare both modelling techniques.

Results

Direct modelling of species richness

All species richness (SR) maps presented in this section (Figure 11; Appendix 5; Appendix 6) were based on species richness observed across the 912 sampling plots (Figure 4).

Predictions by the GLM and the GAM with Poisson distribution were highly correlated to those by the GLM and GAM with zero-inflated negative binomial distribution (spearman rank correlation close to one; Figure 5a, Figure 5b).

Spearman rank correlation between GLM and GAM predictions of SR with the Poisson distribution showed that predictions by the two types of models are highly correlated (93.1%, p-value < 2.2e-16; Figure 6a). The one between the GLM and GAM with the zero-inflated negative binomial shows similar results (92.7%, p-value < 2.2e-16; Figure 6b). As both distributions seem to give quite similar results, only the Poisson distribution is retained in the following results, because this distribution is the most relevant and the most used for counts data.

The fitted GLM – Poisson explained 37.9% of the deviance in SR, while the GAM – Poisson explained 44.8% (Table 3). The GAM also proved better than the GLM across the different evaluation measures (means over 100 runs), (Table 3). Both the mean spearman rank correlation between the observed and predicted SR and the mean ΔAIC were greater for the GAM than for the GLM.

Therefore, in the following, we only present results of the GAM with the Poisson distribution for the direct approach (Figure 7; Figure 11a). All others results and maps obtained with the zero-inflated negative binomial GLM and GAM are given in Appendix 6 and the ones obtained with the Poisson GLM in Appendix 5.

In the Figure 7 is displayed the plot representing the spearman rank correlation of the observed versus predicted SR with the GAM – Poisson. This direct approach predicts the SR around the observed mean.

Cumulative modelling of species richness

AUCs of all models before ensemble forecasting and piling up the predictions, is displayed in Figure 8. For each species, most of models have an AUC higher than 0.7.

Evaluation of piling-up SR predictions was made by a spearman rank correlation between observed and predicted SR (Figure 9). We can see that this cumulative approach tends to overpredict the SR.

Comparing direct versus cumulative modelling

Comparing the pilled-up SR predictions and the direct Poisson GAM yielded a Spearman correlation of 0.803 (p-value < 2.2e-16; Figure 10).

Species richness maps resulting from both approaches – the GAM with the Poisson distribution and the pile of the ensemble forecasting – are displayed in the Figure 11. Despite the fact that both approaches are highly correlated (Figure 10), we can see that the cumulative approach predict globally higher species number than the direct approach. However, the minimum SR is predicted by both approaches at high altitudes and the maximum SR is predicted at intermediate elevations.
Others maps are given in Appendix 5 (GLM – Poisson) and Appendix 6 (GLM and GAM with zero-inflated negative binomial).

Residuals of both approaches plotted as function of the altitude (Figure 12) allows seeing where the models overpredict the species number. The cumulative approach predicts higher SR at every altitude, but this overprediction tends to be more important at low altitude and less marked at high altitude.

The plots of the number of species predicted by each approach as function of the altitude are respectively displayed in the Figure 13a and the Figure 13b. The one for the observed species number is shown in the Figure 13c. The observed SR follows a hump-shaped curve, i.e. less species at high and low elevations and a peak at intermediate altitudes. This hump-shaped curve is more distinct in the direct approach than in the cumulative one. The other particularity of the observed pattern is the high variability in species richness at a same altitude. This pattern is reflected by the cumulative approach and not by the direct one. This characteristic explains the fact that the cumulative approach is better correlated to the reality than the direct approach.

Discussion

Both approaches to modelling species richness – direct and cumulative – yield comparable results and their predictions are highly correlated. However, looking at the distribution of the residuals from both approaches along elevation reveals great differences. Despite the fact that the cumulative approach is better correlated with the observed species richness pattern, it predicts globally too many species, while the direct approach predicts values of species richness around the observed mean. The cumulative approach is biaised. This overprediction by the cumulative approach may come from the fact that the maps obtained by individual SDM use topographic and climatic predictors only and therefore do not take into account all restrictions due for instance to competition between species or land use. The niche fitted by SDMs is thus likely larger than the one observed in nature (true realized

niche). Consequently, pilling potential distribution maps to calculate species richness will overpredict the total number of species occurring in each pixel (Guisan & Rahbek in prep).

This overprediction of species richness by the cumulative approach is detectable when comparing the predicted species richness maps yielded by the two approaches (Figure 11), but is more visible in the plots representing the residuals of both approaches according to altitude (Figure 12). Although large differences between the two approaches can be observed at every altitude, these are greater at lower altitudes, notably in the Plaine du Rhône. One would theoretically expect low altitude areas to shelter more species because smoother climatic conditions prevail and a greater diversity of habitats should be found. However, in the study area, we observed during the field campaign that more human influences also prevail at lower elevations, resulting in more urbanized and productive habitats. Indeed, besides the pure climate influence, plant species richness is thought to be increasingly impacted by anthropogenic activities, such as fire, farming and grazing (Austrheim et al., 1999; Dullinger et al., 2002; Tinner & Theurillat, 2003). For examples, nowadays, many easily accessible areas are highly fertilized and mown while more and more poorly accessible pastures are abandoned (Gellrich & Zimmermann, 2007). These latter are recolonized by forests (Motta & Nola, 2001; Gellrich et al., 2007; Vittoz et al., 2008), inducing the disappearance of many open habitat species. Intensive farming and grazing, characterised by a high level of habitat fertilization and pasturing are increasingly frequent at lower altitudes and can thus have negative impact on plant biodiversity. Indeed, an intensive exploitation as agricultural zones induces the disappearance of natural grasslands and thus indirectly causes the decrease the plant diversity linked to natural habitats. (Luoto et al., 2003). Furthermore, in highly fertilizated regions, only the most competitive species survive and thus supplant the other species (Pausas & Austin, 2001). These phenomena are known to cause important floristic changes, and thus they can have a negative influence on species richness (Walker et al., 2009). The decrease in species richness at low altitudes that is therefore suppose to be linked to human activities, is better pictured by the macroecological approach than by the cumulative one. This is likely because the macroecological model is based on the observed number of species, while the cumulative one consists on summing the potential species distribution. A species could be predicted to be present in low altitude thanks to topographic and climatic predictors, but could also be excluded by others factors that are not included in the models, as land-use in our case. Also, this type of common error is multiplied during the pilling and is therefore causing the difference between the two approaches.

Consequently, testing and verifying the effect of human influences on species richness formally requires taking land-use predictors into account in further modelling study.

Besides the strong overprediction of species richness at low altitude by the cumulative approach, the two species richness prediction maps and graphs of predicted species richness along elevation show overall very similar trends, and both fit reasonably the observed hump-shaped curve trend (Figure 13c; Figure 4). However, the direct approach seems to follow better the observed hump-shape curve than the cumulative approach.

Firstly, the greater the elevation, the smaller the number of species. This trend has already been found in other studies (Stevens, 1992; Austrheim, 2002; Theurillat et al., 2003). Indeed, higher altitudes mean stressful conditions with restricted resources (nutrients), increasingly unstable and steep terrain (screes) and cooler climate, with more frost days. These conditions limit the colonization by plants because they are difficult to endure for a majority of species that are not physiologically adapted to this environment (Körner, 2003). This trend is in agreement with the macroecological hypothesis about the extreme conditions and the lack of resources, and habitat disturbance.

Secondly, as it is shown by several studies, the peak of species diversity occurs in the middle of the elevation gradient (around 1500 m a.s.l.) because these zones are less disturbed by human and agriculture and offer suitable environmental conditions and resources (Austrheim, 2002; Wohlgemuth et al., 2008). More heterogeneous habitats are provided, increasing the diversity of species. Furthermore, it seems that at mid-elevations, plant species of low elevation coexist with plant species of high elevation, resulting in an increase in species richness (Edwards & Armbruster, 1989). Nevertheless, this trend of species diversity peak at mid elevation is more visible in the direct approach, while it is fuzzier in the cumulative approach.

After having discussed the patterns of predictions obtained with each of the two modelling approaches, it is now interesting to discuss their respective strengths and weaknesses.

The direct approach predicts the total number of species that can potentially be found in a geographic unit, but looses the information about which species are occurring there. As a result, two sites with very different environmental features can have the same number of species but not the same species composition. Climatic and topographic variables used in the modelling therefore determine the total number of species present in a unit without distinguishing which species coexist. This may be a problem when forecasting the potential impact of climate change on plant biodiversity. Indeed, a site under global warming may well exhibit a different species composition while its total number of species remains the same.

In contrast, the prediction patterns resulting from the cumulative approach take into account which species co-occur in each unit. However, the cumulative approach suffers from intrinsic limitations to SDMs. Firstly, SDMs assume that each modelled species is in equilibrium with its environment. That is, a species is expected to have colonized all regions where it could

potentially grow. Climatic and topographic predictors used in the model determine the envelope of occurrences of each species. This is the reason why this approach overpredicts the number of species. Indeed, it forecasts species in all regions that have similar environmental characteristics to their current range, even in regions that cannot be colonized in the reality by some of the species. That is not the case with the direct approach, which does not predict each species separately, but models a number of species. As, with the cumulative technique a habitat suitability map for each species is obtained and then all maps are pilled up, the error accumulates and thus decreases the general relevance of the model. The second limitation concerns rare species (Guisan & Rahbek, in prep) that are difficult to model in SDMs (Guisan & Thuiller, 2005). Rare species were not a problem in this study, because only the species that were observed more that 20 times throughout all sampling plots were selected (minimum prevalence). However, only modelling the most frequent species may not work for addressing conservation biology issues, because often the rarest species are also the ones needing greatest protection and management (e.g. through models). Hence, they may represent differences in species richness that one want to evidence with models.

A solution to modelling biodiversity that combines the advantages of both approaches while resolving some of their problems would be to overpass the limitations of the cumulative approach overpredicting species richness with constraints defined by the macroecological approach, as proposed in Guisan & Rahbek (in prep.), so to combine both approaches into one single unified framework. This new framework consists in using macroecological models for defining the maximum number of species potentially occurring in a unit, and modelling independently the distribution of all species in the local pool. Then the final plant species assemblage is predicted by applying successive filters (e.g. dispersal) or assembly rules on the predicted species pool to reach the maximum number of species previously defined by the macroecological model.

Summarizing, in this study, I observed that both approaches yield quite similar spatial patterns of species richness, but have both strengths and weakness. The cumulative approach is based on the prediction of each species separately and therefore tends to overpredict species richness, mostly at lower altitude, whereas the direct approach gives only a number of species that is rather close to the one observed in nature. Thus, it seems that both approaches should rather be seen as complementary. As suggested by Guisan & Rahbek (in prep.), combining both approaches into a single unifying framework may be promising for modelling biodiversity because it could reduce prediction mismatches.

Acknowledgments

This study was performed in the Spatial Ecology Group in the Department of Ecology and Evolution of the University of Lausanne and was supervised by Antoine Guisan and Anne Dubuis. I would like specially to thank Anne Dubuis and Antoine Guisan for their precious technical and theoretical advices concerning the modelling and the corrections on my manuscript. I would also like to thank Julien Pottier, Loïc Pellisier and Blaise Petitpierre that gave me technical support for the modelling. Finally, I am very grateful to all people that helped for the data collection on the field during the summer.

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. – In: Petrov, B.N. and Csaki, F. (eds), Second International Symposium on information theory. Akademiai Kiado, Budapest, pp. 267 – 281.

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**, pp. 22 – 33.

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

Austin, M.P. (1999) The potential contribution of vegetation ecology to biodiversity research. *Ecography*, **22**, pp. 465 – 484.

Austrheim, G., Gunilla, E., Olsson, A. & Grøntvedt, E. (1999) Land-use impact on plant communities in semi-natural sub-alpine grasslands of Budalen, central Norway. *Biological Conservation*, **87**(3), pp. 369 – 379.

Austrheim, G. (2002) Plant diversity patterns in semi-natural grasslands along an elevational gradient in southern Norway. *Plant Ecology*, **161**, pp. 193 – 205.

Beniston, M. (2006) Mountain weather and climate: a general overview and a focus on climatic change in the Alps. *Hydrobiologia*, **562**, pp. 3 – 16.

Bergamini, A., Stofer, S., Bolliger, J. & Scheidegger, C. (2007) Evaluationg macrolichens and environmental variables as predictors of the diversity of epiphytic microlichens. *The Lichenologist*, **39 (5)**, pp. 475 – 489.

Bouët, M. (1985) Climat et météorologie de la Suisse romande. Payot, Lausanne, Switzerland, 170 p.

Braun-Blanquet, J. (1964) Pflanzensoziologie. Grundzüge der Vegetationskunde, 3rd edition. Springer Verlag, Wien/New York, 865 p.

Connell, J.H., (1978) Diversity in tropical rainforests and coral reefs. *Science*, **199**, pp. 1302 – 1310.

Currie, D.J. (1991) Energy and large-scale patterns of animal- and plant- species richness. *The American Naturalist*, **137**, pp. 27 – 49.

Diaz, H.F., Grosjean, M. & Graumlich, L. (2003) Climate variability and change in high elevation regions: past, present and future. *Climatic Change*, **59**, pp. 1 – 4.

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

558 Dobbertin, M., Hilker, N., Rebetez, M., Zimmermann, N.E., Wohlgemuth, T. & Rigling, A.
559 (2005) The upwart shift of pine mistletoe (Viscum album ssp. austriacum) in Switzerland –
560 the result of climate warming? *International Journal of Biometeorology*, **50**, pp. 40 – 47.

561 Drake, J.M., Randin, C. & Guisan, A. (2006) Modelling ecological niches with support
562 vector machnes. *Journal of Applied Ecology*, **43**, pp. 424 – 432.

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

566 Dullinger, S., Dirnböck, T. & Grabherr, G. (2002) Effects of summer farming on
567 subalpine plant species diversity. *Proceedings 32nd Annual conference of the Ecological*
568 *Society of Germany, Austria and Switzerland*. Cottbus, Germany, 2002, 16. - 20 September.

569 Edwards, M.E. & Armbruster, W.S. (1989) A tundra-steppe transition on Kathul
570 mountain, Alaska, U.S.A. *Arctic and Alpine Research*, **21(3)**, pp. 296 – 304.

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

574 Engler, R., Randin, C., Thuiller, W., Dullinger, S., Zimmermann, N.E., Araújo, M.B.,
575 Pearman, P.B., Albert, C.H., Choler, P., de Lamo, X., Dirnböck, T., Gómez-García, D.,
576 Grytnes, J.A., Heegard, E., Høistad, F., Le Lay, G.,Nogués-Bravo, D., Normand, S., Piédalu,
577 C., Puscas, M., Sebastià, M.T., Stanisci, A., Theurillat, J.P., Trivedi, M., Vittoz, P. & Guisan ,
578 A. (in prep.) Climate change impacts on European mountain plant diversity.

579 ESRI (2008) ARCInfo Version 9.2. Environmental Systems Research Institute, Inc.,
580 Redlands, CA, US.

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

583 Fisher, R.A., Corbet, S. & Williams, C.B. (1943) A theoretical distribution for the apparent
584 abundance of different species. *Journal of Animal Ecology*, **12 (1)**, pp. 54 – 58.

585 Folke, C., Holling, C.S. & Perrings, C. (1996) Biological diversity, ecosystems, and the
586 human scale. *Ecological Applications*, **6 (4),** pp. 1018 – 1024.

587 Gehrig-Fasel, J., Guisan, A. & Zimmermann, N.E. (2007) Tree line shifts in the Swiss
588 Alps: climate changes or land abandonment? *Journal of Vegetation Science,* **18**, pp. 571 –
589 582.

590 Gellrich, M. & Zimmermann, N.E. (2007) Investigating the regional-scale pattern of
591 agricultural land abandonment in the Swiss mountains: A spatial statistical modelling
592 approach. *Landscape and Urban Planning,* **79(1)**, pp. 65 – 76.

593 Gellrich, M., Baur, P. & Zimmermann, N.E. (2007) Natural forest regrowth as a proxy
594 variable for agricultural land abandonment in the Swiss mountains: a spatial statistical model
595 based on geophysical and socio-economic variables*. Environmental Modeling & Assessment*
596 **12(4)**, pp. 269 – 278.

597 Gotelli, N.J., Anderson, M.J., Arita, H.T., Chao, A., Colwell, R.K., Connolly, S.R., Currie,
598 D.J., Dunn, R.R., Graves, G.R., Green, J.L., Grytnes, J.-A., Jiang, Y.-H., Jetz, W., Lyons,
599 S.K., McCain, C.M., Magurran, A.E., Rahbek, C., Rangel, T.F.L.V.B., Webb, C.O. & Willig,
600 M.R. (2009) Patterns and causes of species richness: a general simulation model for
601 macroecology. *Ecology Letters*, **12**, pp. 873 – 886.

602 Guisan, A., Holten, J.I., Spichiger, R. & Tessier, L. (1995) Potential Ecological Impacts of
603 Climate Change in the Alps and Fennoscandian Mountains. Conservatoire et Jardin
604 Botaniques, Geneva, Switzerland, 194 p.

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

607 Guisan, A. & Theurillat, J.P. (2000b) Assessing alpine plant vulnerability to climate
608 change: a modeling perspective. *Integrated Assessment*, **1**, pp. 307 – 320.

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

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

613 Guisan, A & Rahbek, C. (in prep) Predicting spatio-temporal patterns of species
614 assemblages through integration of macroecological and species distribution models with
615 assembly rules and source pool assignments.

616 Gutiérrez Teira, A. & Peco, B. (2003) Modelling oldfield species richness in a mountain
617 area. *Plant Ecology*, **166**, pp. 249 – 261.

618 Hastie, T. & Tibshirani, R. (1990) Generalized additive models. Chapman & Hall,
619 London, 356 p.

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

622 Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbour Symposium on*
623 *Quantitative Biology*, **22**, pp. 415 – 427.

624 Kissling, W.D., Rahbek, C. & Böhning – Gaese, K. (2007) Food plant diversity as broad-
625 scale determinant of avian frugivore richness. *Proceedings of The Royal Society*, **274**, pp.
626 799 – 808.

627 Körner, C. (2003) Alpine plant life: functional plant ecology of high mountain
628 ecosystems. Springer-Verlag Berlin Heidelberg, 359 p.

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

632 Lenihan, J.M. (1993) Ecological response surfaces for north American boreal tree
633 species and their use in forest classification. *Journal of Vegetation Science*, **4 (5)**, pp. 667 –
634 680.

635 Lobo, J.M. & Martín-Piera, F. (2002) Searching for a predictive model for a species
636 richness of Iberian dung beetle based on spatial and environmental variables. *Conservation*
637 *Biology*, **16 (1)**, pp. 158 – 173.

638 Lobo, J.M., Jay-Robert, P. & Lumaret, J.P. (2004) Modelling the species richness
639 distributions for French Aphodiidae (Coleoptera, Scarabaeoidea). *Ecography*, **27**, pp. 145 –
640 156.

641 Luoto, M., Rekolainen, S.P., Aakkula, J. & Pykalä, J. (2003) Loss of plant species
642 richness and habitat connectivity in grasslands associated with agricultural change in
643 Finland. *Ambio*, **32(7)**, pp. 447 – 452.

644 Maggini R., Guisan, A. & Cherix, D. (2002) A stratified approach for modelling the
645 distribution of a threatened ant species in the Swiss National Park. *Biodiversity and*
646 *Conservation,* **11**(12), pp. 2117 – 2141.

647 Margules, C.R. (1989) Introduction to some Australian developments in conservation
648 evaluation. *Biological Conservation*, **50**, pp. 1 – 11.

649 Marini, L., Prosser, F., Klimek, S. & Marrs, R.H. (2008) Water-energy, land-cover and
650 heterogeneity drivers of the distribution of plant species richness in mountain region of the
651 European Alps. *Journal of Biogeography*, **35**, pp. 1826 – 1839.

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

655 McCullagh, P. & Nedler, J.A. (1989) Generalized linear models. 2nd edition. Chapman &
656 Hall, London, UK, 532 p.

Motta, R. & Nola, P. (2001) Growth trends and dynamics in subalpine forest stands in the Varaita Valley (Piedmont, Italy) and their relationships with human activities and global change. *Journal of Vegetation Science*, **12**, pp. 219 – 230.

Nogués-Bravo, D., Araújo, M.B., Martinez-Rica, J.P. & Errea, M.P. (2007) Exposure of global mountain systems to climate warming during the 21st century. *Global Environmental Change*, **17 (3-4)**, pp. 420 – 428.

Olden, J.D. (2003) A species-specific approach to modelling biological communities and its potential for conservation. *Conservation Biology,* **17**, pp. 854 – 863.

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(11)**, pp. 2501 – 2509.

Pauli, H., Gottfried, M. & Grabbherr, G. (1996) Effects of climate change on mountain ecosystems. Upward shifting of alpine plants. *World Resource Review*, **8**, pp. 382 – 390.

Pauli, H., Gottfried, M., Reiter, K., Klettner, C. & Grabherr, G. (2007) Signals of range expansions and contractions of vascular plants in the high Alps: observations (1994–2004) at the GLORIA master site Schrankogel, Tyrol, Austria. *Global Change Biology*, **13**, pp. 147 – 156.

Pausas, J.G. & Austin, M.P. (2001) Patterns of plant species richness in relation to different environments: an appraisal. *Journal of Vegetation Science*, **12 (2)**, pp. 153 – 166.

Pearson, R.G., Dawson, T.P., Berry, P.M. & Harrison, P.A. (2002) SPECIES: a spatial evaluation of climate impact on the envelope of species. *Ecological Modelling*, **154**, pp. 28 – 300.

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

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org.

Rabhek, C., Gotelli, N.J., Colwell, R.K., Entsminger, G.L., Rangel, T.F.L.V.B. & Graves, G.R. (2007) Predicting continental-scale patterns of bird species richness with spatially explicit models. *Proceedings of The Royal Society*, **274**, pp. 165 – 174.

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

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

694 Randin, C.F., Jaccard, H., Vittoz, P., Yoccoz, N.G. & Guisan, A. (2009b) Land use
695 improves spatial predictions of mountain plant abundance but not presence-absence*. Journal*
696 *of Vegetation Science*, **20 (6)**, pp. 996 – 1008.

697 Saetersdal, M., Birks, H.J.B. & Peglar, S.M. (1998) Predicting changes in
698 Fennoscandian vascular-plant species richness as a result of future climatic change. *Journal*
699 *of Biogeography*, **25**, pp. 111 – 122.

700 Schmidt, M., König, K. & Müller, J.V. (2008) Modelling species richness and life form
701 compostion in Sahelian Burkina Faso with remote sensing data. *Journal of Arid*
702 *Environments*, **72**, pp. 1506 – 1517.

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

705 Stevens, G.C. (1992) The elevational gradient in altitudinal range: an extension of
706 Rapoport's latitudinal rule to altitude. *The American Naturalist*, **140**, pp. 893 – 411.

707 Swets, J.A. (1988) Measuring the Accuracy of Diagnostic Systems. *Science*, **240 (4857),**
708 pp. 1285 – 1293.

709 Theurillat, J.P., Felber, F., Geissler, P., Gobat, J.M., Fierz, M., Fischlin, A., Küpfer, P.,
710 Schlüssel, A., Velluti, C., Zhao, G.F & Williams, J. (1998) Sensitivity of plant and soil
711 ecosystems of the Alps to climate change. In: Views from the Alps. Regional Perspectives on
712 Climate Change (eds Cebon P, Dahinden U, Davies HC, Imboden D, Jäger CC), MIT Press,
713 London, UK, pp. 225 – 308.

714 Theurillat, J.P., Schlüssel, A., Geissler, P., Guisan, A., Velluti, C. & Wiget, L., (2003)
715 Vascular plant and bryophyte diversity along elevation gradients in the Alps (chapter 8). In
716 Alpine Biodiversity in Europe L. Nagy, G. Grabherr, C. Körner and D. B. A. Thompson.
717 Berlin. *Ecological Studies*, **167**, pp. 185 – 194.

718 Thuiller, W. (2003) BIOMOD – Optimizing predictions of species distributions and
719 projecting potential future shifts under global change. *Global Change Biology*, **9**, pp. 1353 –
720 1362.

721 Thuiller, W., Lavorel, S., Araujo, M.B., Sykes, M.T. & Prentice, I.C. (2005) Climate
722 change threats to plant diversity in Europe. *Proceedings of the National Academy of Science*,
723 **102 (23)**, pp. 8245 – 8250.

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

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

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009b) ModOperating Manual for BIOMOD, 90 p.

Tilman, D. (1996) Biodiversity: population versus ecosystem stability. *Ecology*, **77**, pp. 350 – 363.

Tinner W. & Theurillat, J.P. (2003) Uppermost limit, extent, and fluctuations of the timberline and treeline ecocline in the Swiss Central Alps during the past 11,500 years. Arctic, *Antarctic and Alpine Research*, **35**, pp. 158 – 169.

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

Vittoz, P., Jutzeler, S. & Guisan, A. (2006) Flore alpine et réchauffement climatique: observation de trois sommets valaisans à travers le 20ème siècle. *Bulletin de la Murithienne*, **123**, pp. 49 – 59.

Vittoz, P., Rulence, B., Largey, T. & Freléchoux, F. (2008) Effects of climate and land-use change on the establishment and growth of cembran pine (Pinus cembra L.) over the altitudinal treeline ecotone in the Central Swiss Alps. *Arctic Antarctic and Alpine Research*, **40**, pp. 225 – 232.

Walker, K.J., Preston, C.P. & Boon, C.R. (2009) Fifty years of change in area of intensive agriculture: plant trait responses to habitat modification and conservation, Bedfordshire, England. *Biodiversity Conservation*, **18**, pp. 3597 – 3613.

Walther, G.R. (2002) Weakening of climatic constraints with global warming and its consequences for evergreen broadleaved species*. Folia Geobotanica,* **37**, pp. 129 – 139.

Walther, G.R. (2003) Plants in a warmer world. *Perspectives in Plant Ecology, Evolution and Systematics*, **6**, pp. 169 – 185.

Walther, G.R., Beiner, S. & Burga, C.-A. (2005a) Trends in the upward shift of alpine plants. *Journal of Vegetation Science*, **16**, pp. 541 – 548.

Walther, G.R., Berger, S. & Sykes, M.T. (2005b) An ecological ‘‘footprint’’ of climate change. *Proceedings of the Royal Society B*, **272**, pp. 1427 – 1432.

Welsh, A.H., Cunningham, R.B., Donnelly, C.F. & Lindenmayer, D.B. (1996) Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, **88**, pp. 297 – 308.

Whittaker, R.J., Willis, K.J. & Field, R. (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, **28**, pp. 453 – 470.

Wohlgemuth, T., Nobis, M.P., Kienast, F. & Plattner, M. (2008) Modelling vascular plant diversity at the landscape scale using systematic samples. *Journal of Biogeography*, **35**, pp. 1226 – 1240.

Yachi, S. & Loreau, M. (1999) Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proceedings of the National Academy of Sciences*, **96**, pp. 1463 – 1468.

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

Zimmermann, N.E., Yoccoz, N.G., Edwards, T.C., Meier, E.S., Thuiller, W., Guisan, A., Schmatz, D.R. & Pearman, P.B. (2009) Climatic extremes improve predictions of spatial patterns of tree species. *Proceedings of the National Academy of Science (USA),* **106**, pp. 19723 – 19728.

Tables

Table 1. Estimation of plant abundance according to Braun-Blanquet (1964) modified scale used in this study.

Braun-Blanquet scale	Estimation of plant abundance (%)
r	1-2 individuals
+	>1%
1	1-5%
2a	6-13%
2b	14-25%
3	26-50%
4	51-75%
5	>75%

Table 2. Climatic and topographic variables used in this study.

Variables	Abbreviations	Units	Details	References
Degree-days	fddeg0	°C * day * year^{-1}	sum of days multiplied by the daily mean temperature higher than the threshold (0°C)	Zimmermann and Kienast (1999)
Moisture index (average of values of June-August months)	fmmind68	mm * day^{-1}	monthly average of water availability (precipitation - evapotranspiration)	Zimmermann and Kienast (1999)
Global solar radiations (sum for all the year)	fsumradyy	kJ * m^{-2} * year^{-1}	sum of monthly average of daily global potential shortwave radiation	Zimmermann et al. (2009)
Topographic position	ftopos	-	concavity or convexity of the terrain	Zimmermann et al. (2009)
Slope	fslope	degrees	incline of the terrain	ArcGIS (ESRI, 2008)

Table 3. Mean evaluation measures obtained from the repeated split-sample (N=100) evaluation of the GLM and GAM species richness models.

Models	mean spearman rank correlation	mean ΔAIC	mean D^2	mean D^2 adjusted
GLM-Poisson	0.5580654	406.218	0.3778826	0.379505
GAM-Poisson	0.5865458	833.9438	0.4465942	0.4485197

Figures

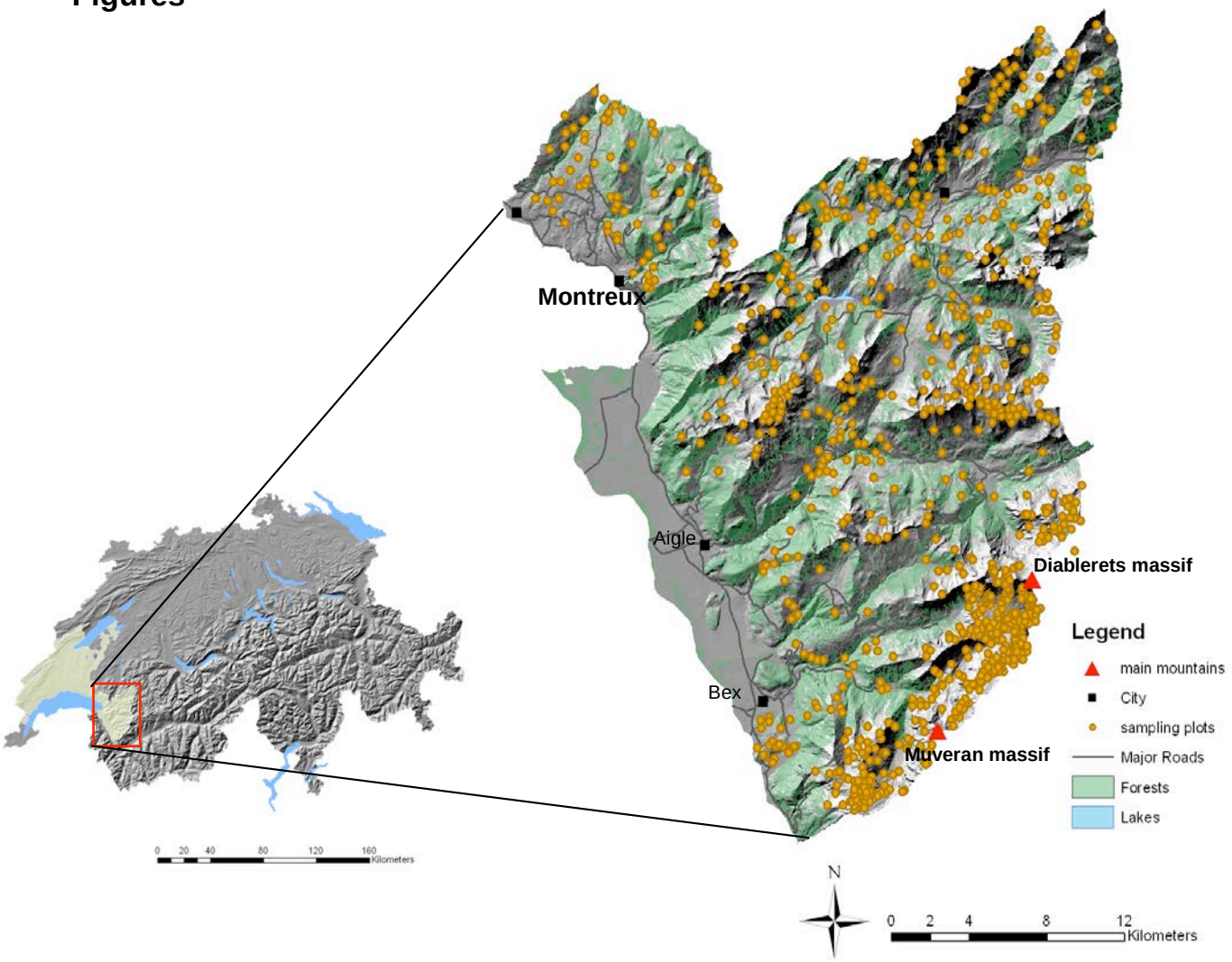

Figure 1. Geographical location of the study area in Switzerland (Préalpes vaudoises), with the different sampling plots in orange.

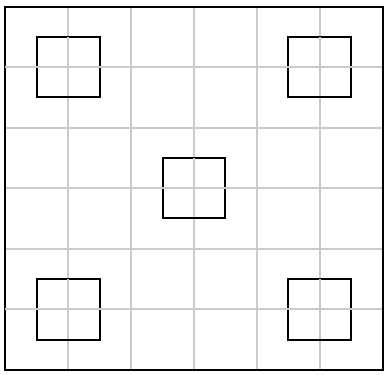

Figure 2. Schema of the 4 m^2 grid used in the study, with the five subplots (20 cm x 20 cm) on each angle and on the centre of the grid.

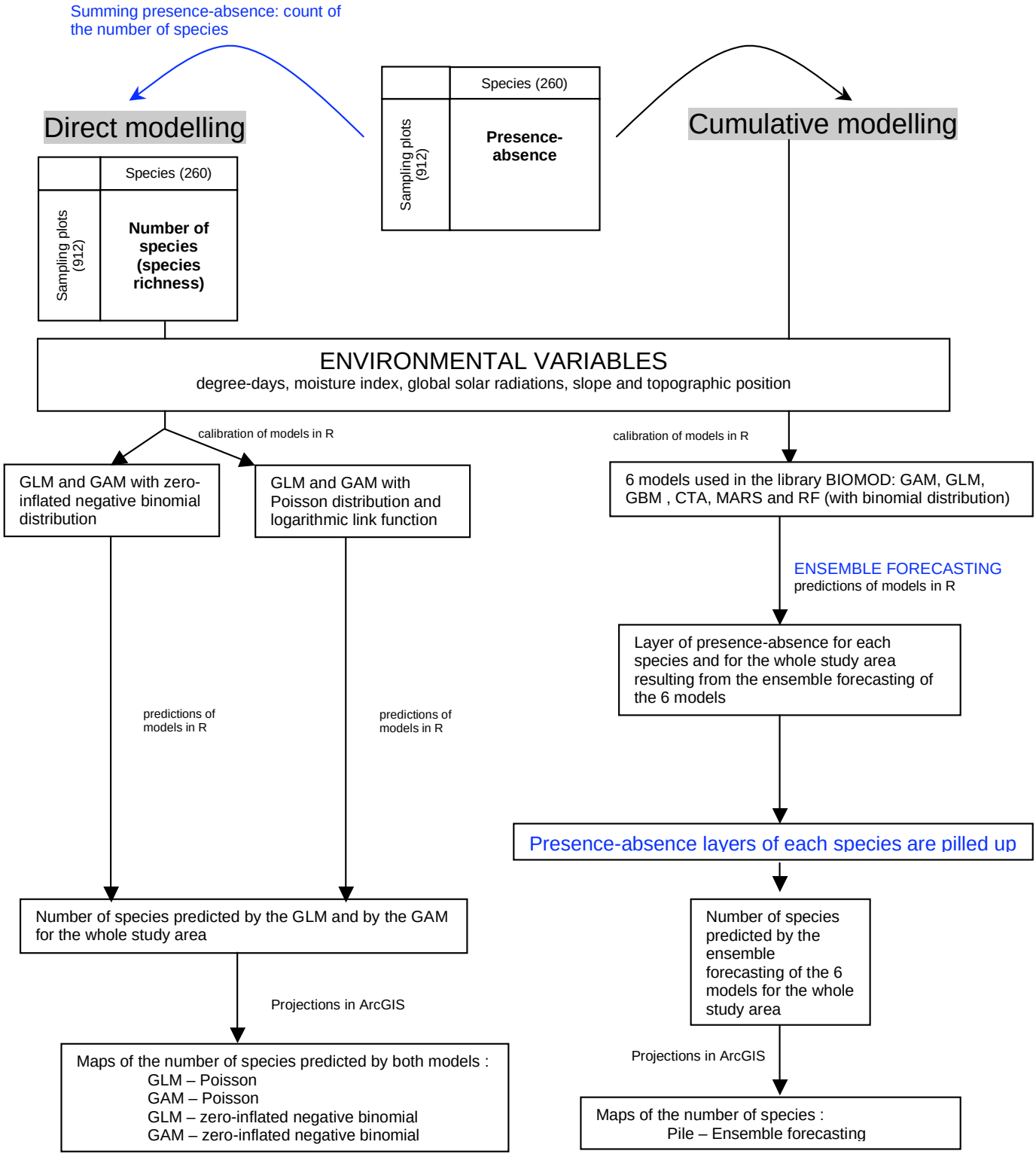

Figure 3. Schema of the two separate approaches used in this study to model species richness (see text) and comparison of their spatial predictions.

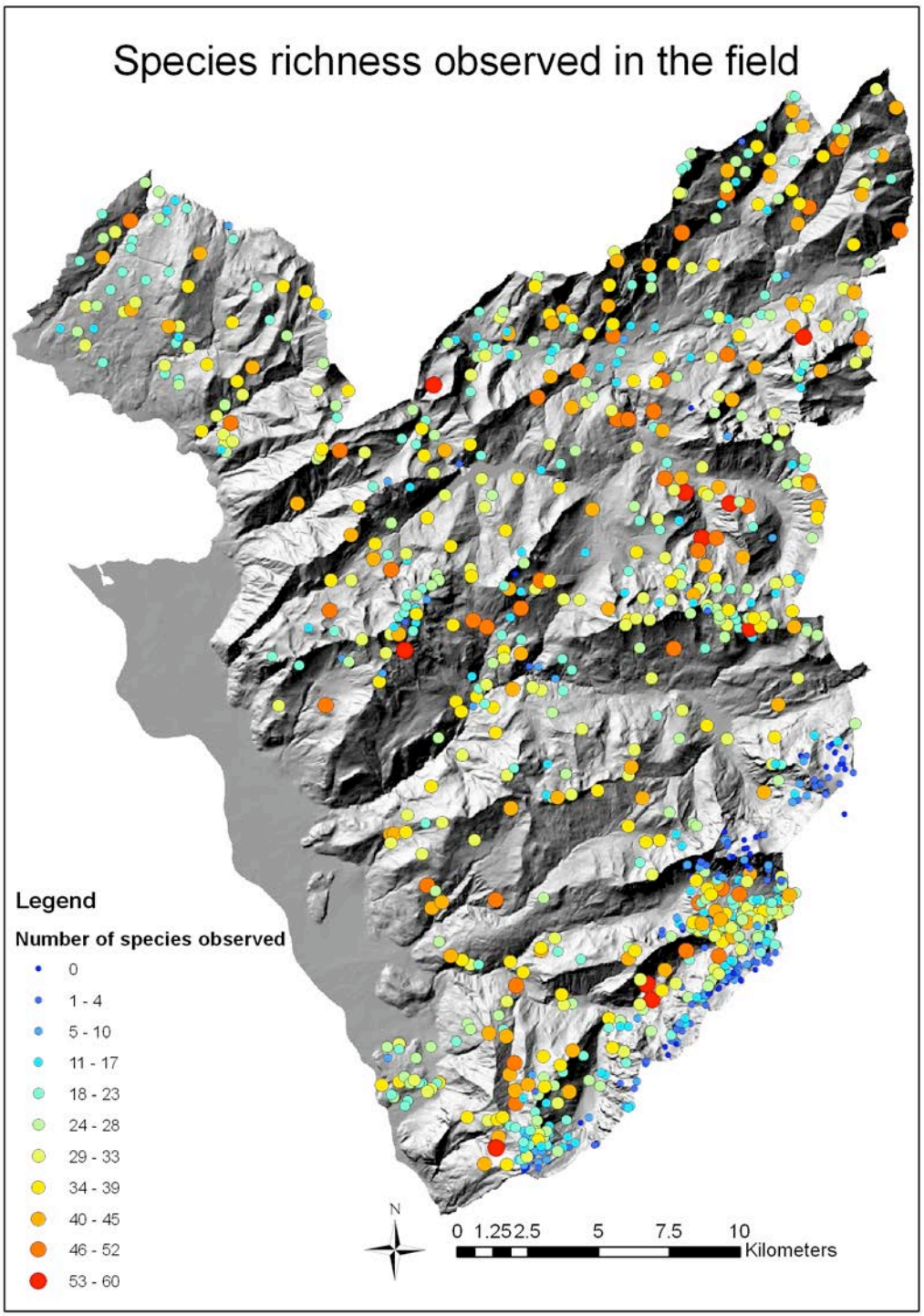

Figure 4. Map of species richness observed across the 912 sampling plots. In blue are the plots with the smallest number of species, whereas in red are those with the highest species richness.

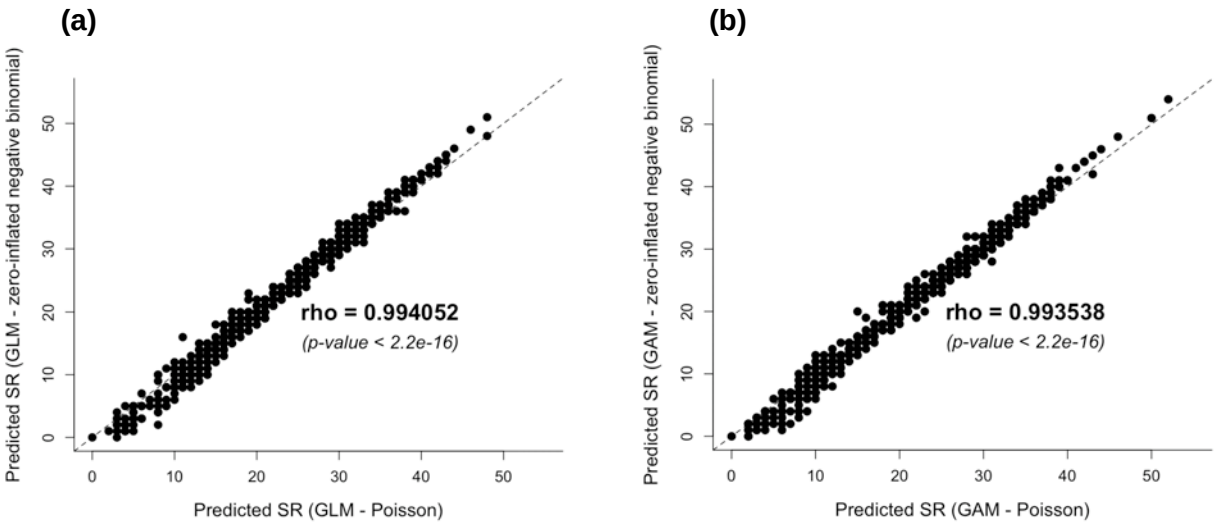

Figure 5. (a) Scatterplot and spearman rank correlation between SR predicted by the GLM with zero-inflated negative binomial distribution and SR predicted by the GLM with Poisson distribution. (b) The same for the GAM model. The dotted line (--) corresponds to a correlation of 1 between the SR predicted by both distributions.

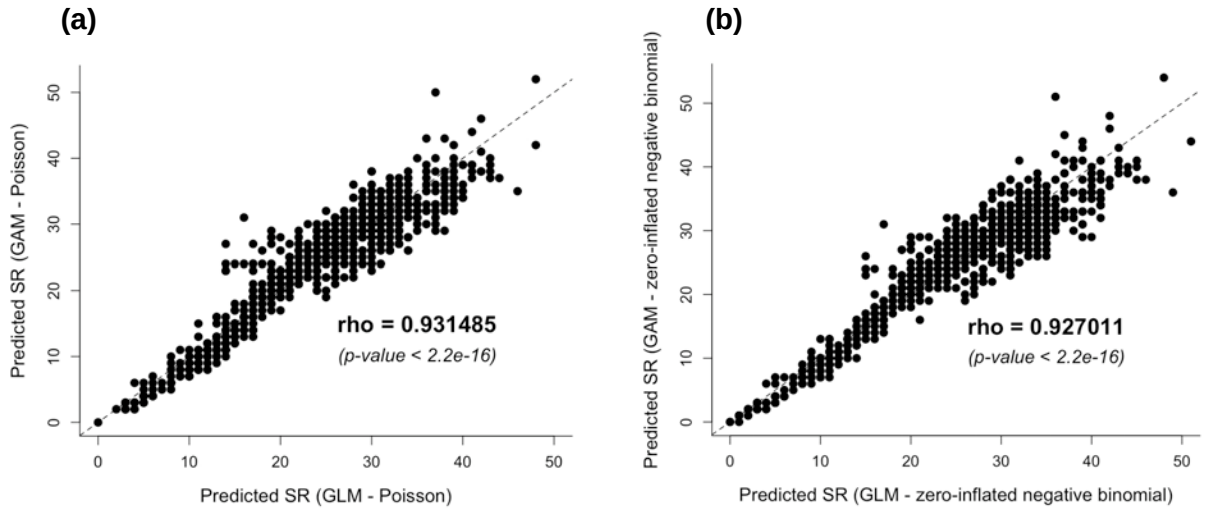

Figure 6. Spearman correlation between GLM and GAM predictions of SR, for (a) the Poisson distribution and (b) the zero-inflated negative binomial distribution. The dotted line corresponds to a perfect correlation between the SR predicted by both models.

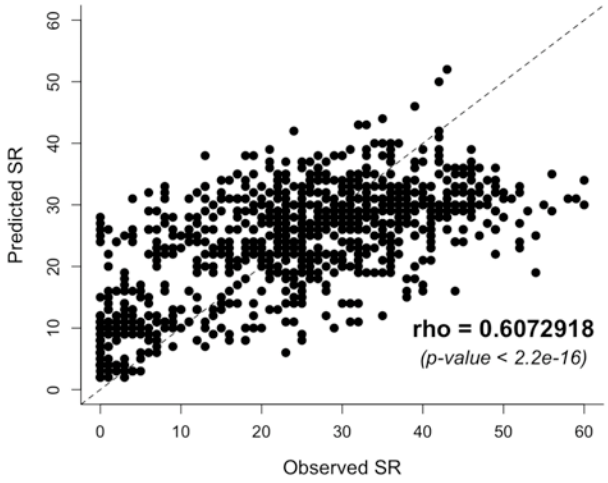

Figure 7. Scatterplot of observed versus predicted SR. The latter were obtained here by the GAM with the Poisson distribution (direct approach). The dotted line corresponds to a correlation of 1 between observed and predicted SR.

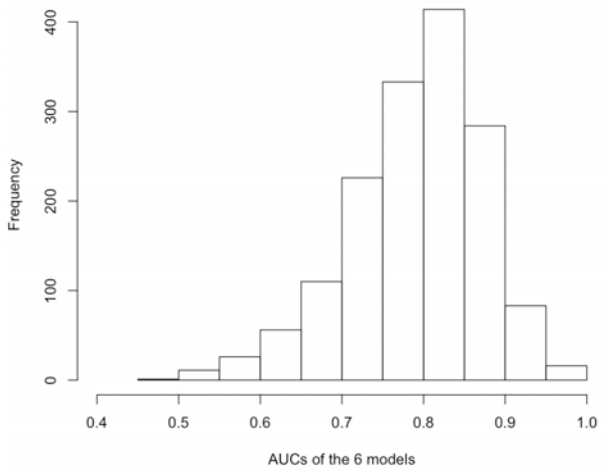

Figure 8. Histogram of the AUCs of the six models used in this cumulative approach.

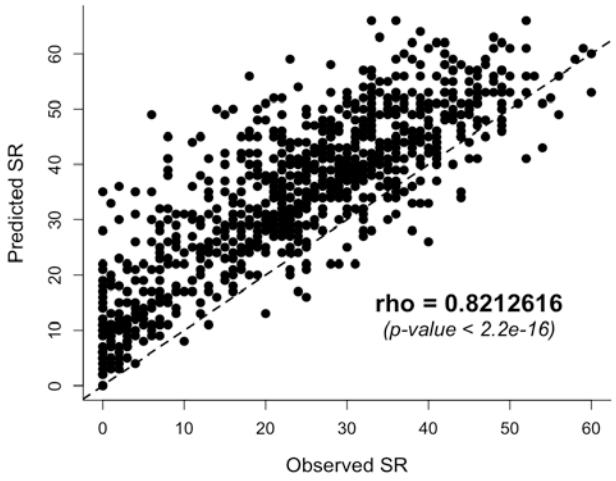

Figure 9. Scatterplot of observed versus predicted SR. The latter were obtained here by piling up ensemble predictions of species distributions (cumulative approach). The dotted line corresponds to a correlation of 1 between observed and predicted SR.

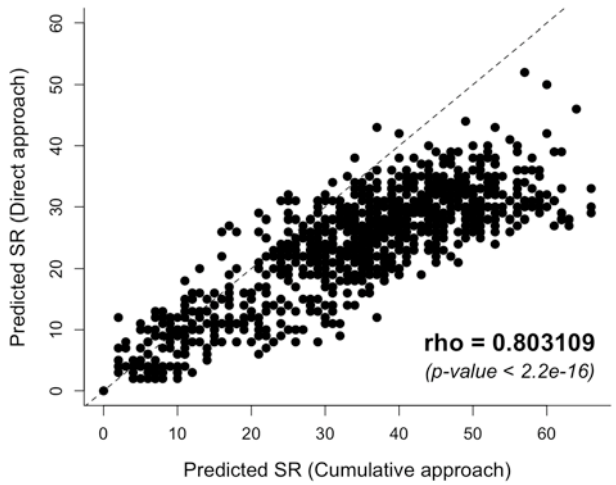

Figure 10. Plots of the spearman correlation between the species richness predicted by the pile of the ensemble forecasting (cumulative approach) and by the GAM with the Poisson distribution (direct approach). The dotted line corresponds to a correlation of 1 between the SR predicted by both approaches.

(a) **(b)**

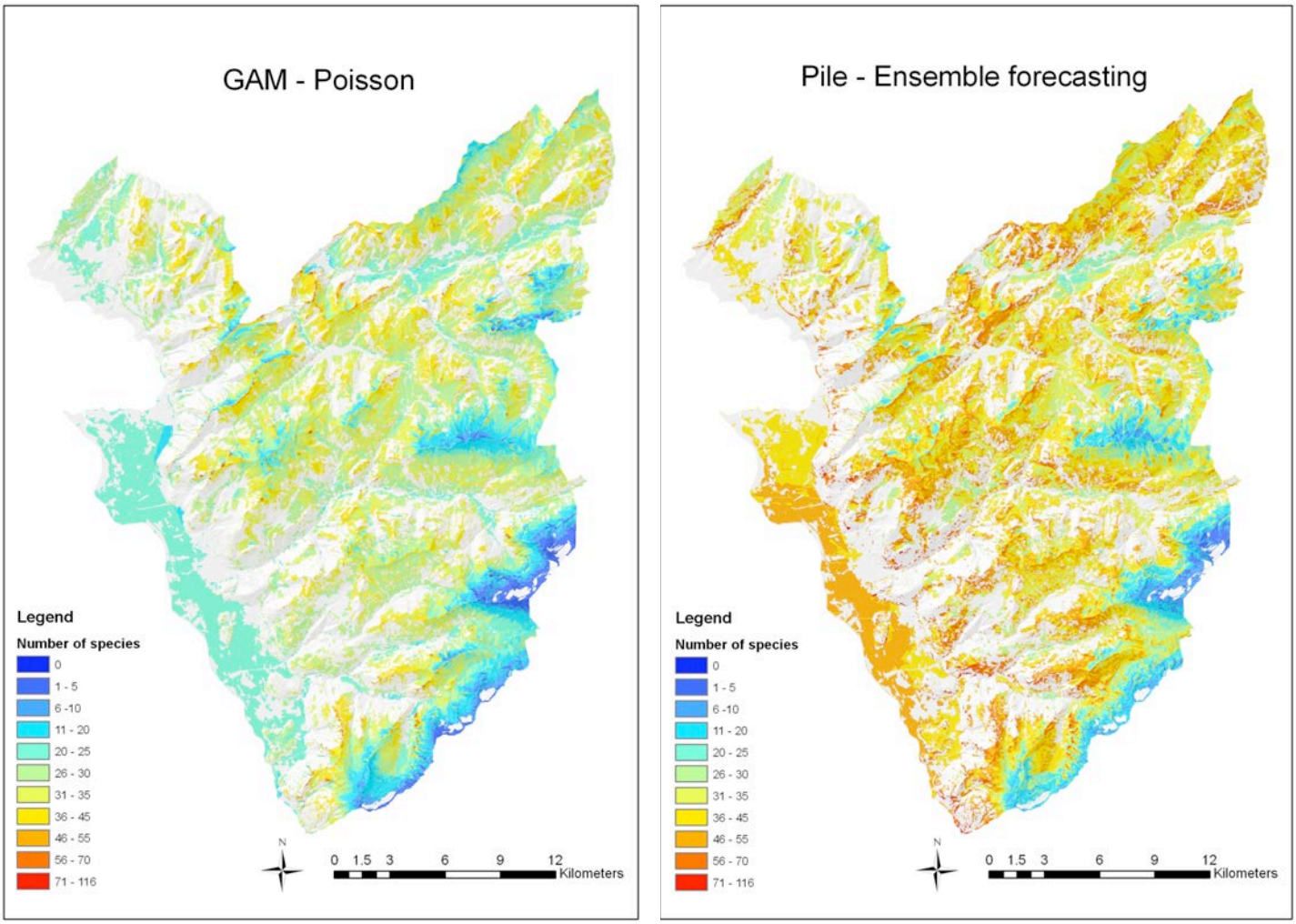

Figure 11. Spatial distribution of the species richness, resulting from (a) the GAM with the Poisson distribution (direct approach) and (b) the pile of the 260 presence-absence ensemble forecasting layers (cumulative approach). Regions in white represent areas that were not used for the projections (glaciers, urbanized zones, lakes and forests). Zones in blue and in red correspond respectively to areas with the smallest and the highest number of species.

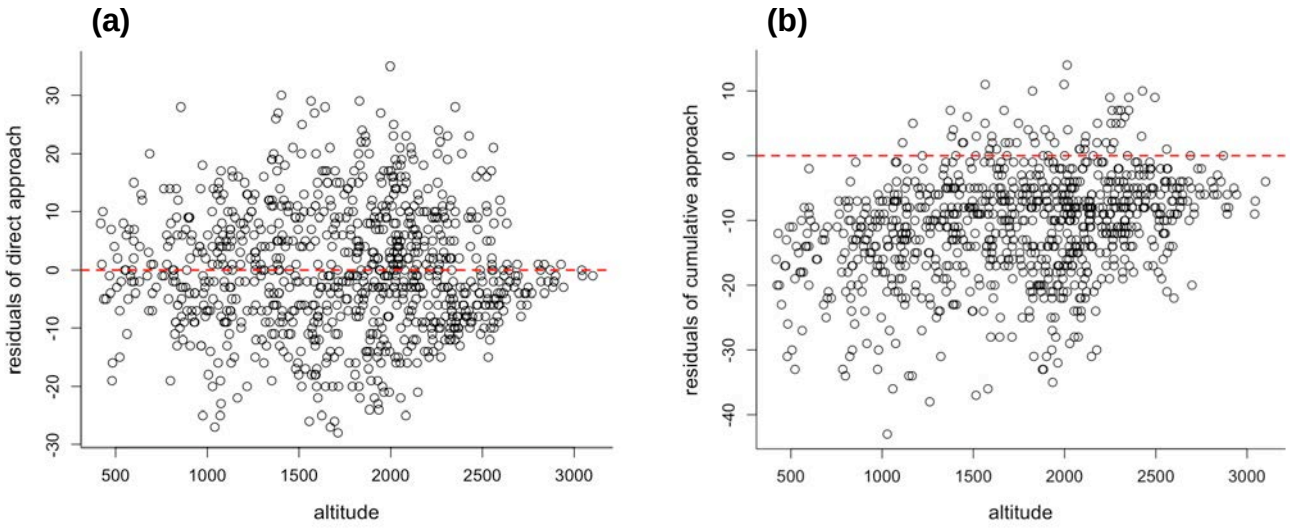

Figure 12. Plots of the residuals of the (a) GAM – Poisson (direct approach) and (b) pilling (cumulative approach). These errors of predictions are plotted as function of the altitude. Values below 0 mean that the predicted species number is higher than the number of species observed (overprediction) and values above 0 mean that predicted species number is lower than the observed one (underprediction). The red dotted line represents the values of 0 corresponding to an equal number of predicted and observed species.

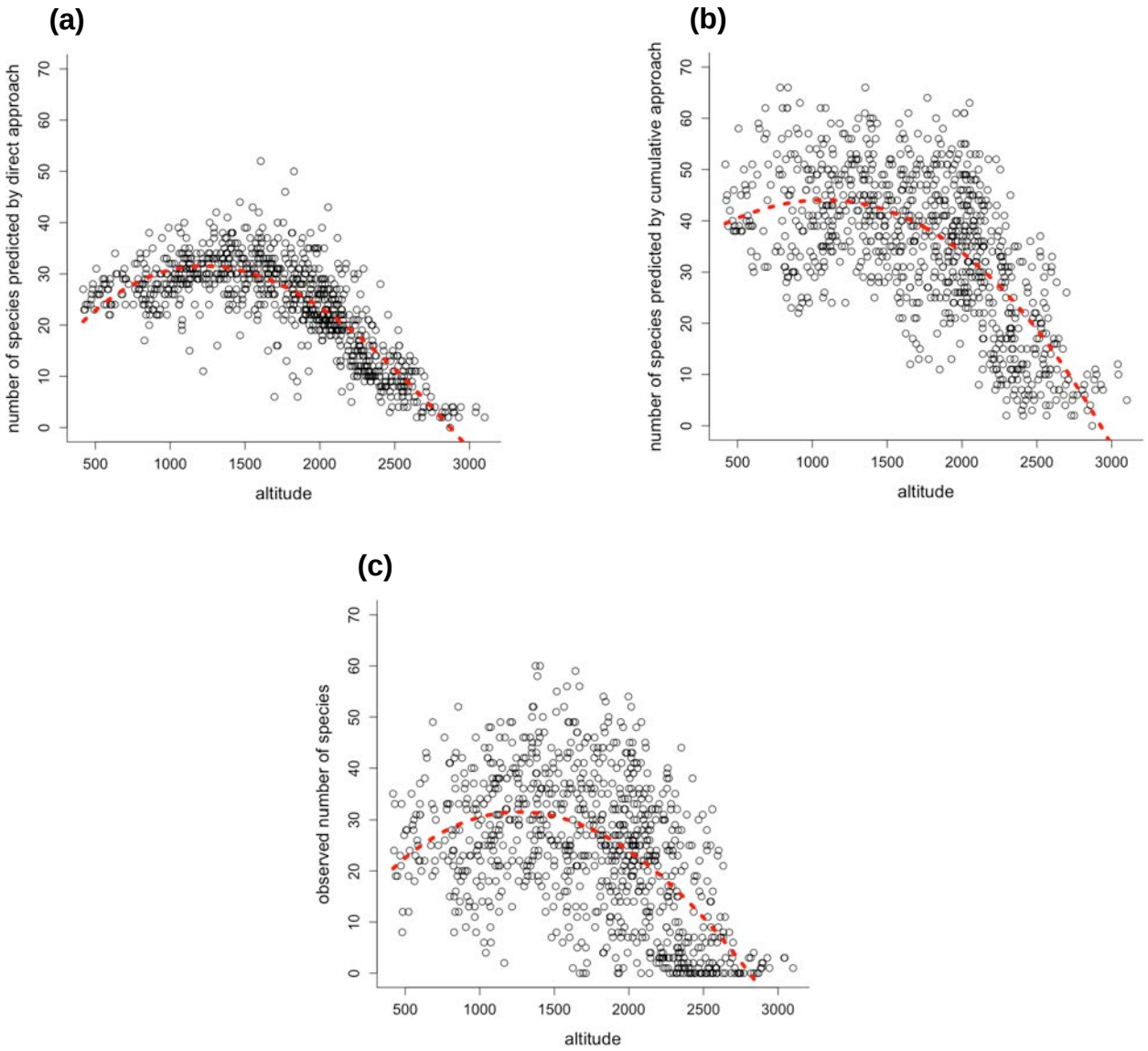

Figure 13. Plots of the number of species predicted by (a) the GAM – Poisson (direct approach) and (b) the pilling of ensemble forecasting maps (cumulative approach), as function of the altitude. The plot of the observed species number as function of the altitude is displayed in (c). Trends are shown by fitting cubic functions (red dotted curve).

APPENDIX 1: List of the 260 plant species used in this study

Table 1A. List of the 260 plant species used in this study. These species were present more than 20 times over all the sampling plots.

Acer pseudoplatanus	Bromus erectus s.str	Cirsium palustre	Galium pumilum
Achillea atrata	Bromus hordeaceus	Cirsium spinosissimum	Gentiana acaulis
Achillea millefolium	Calamagrostis varia	Clinopodium vulgare	Gentiana bavarica
Acinos alpinus	Caltha palustris	Colchicum autumnale	Gentiana campestris s.str
Adenostyles glabra	Campanula barbata	Crepis aurea	Gentiana clusii
Agrostis alpina	Campanula cochleariifolia	Crepis biennis	Gentiana lutea
Agrostis capillaris	Campanula rhomboidalis	Crepis pyrenaica	Gentiana nivalis
Agrostis rupestris	Campanula rotundifolia	Crocus albiflorus	Gentiana purpurea
Agrostis schraderiana	Campanula scheuchzeri	Cruciata laevipes	Gentiana verna
Agrostis stolonifera	Cardamine pratensis	Cynosurus cristatus	Geranium sylvaticum
Ajuga reptans	Carduus defloratus s.l.	Dactylis glomerata	Geum montanum
Alchemilla conjuncta aggr.	Carex atrata aggr.	Dactylorhiza fuchsii	Geum rivale
Alchemilla coriacea aggr.	Carex caryophyllea	Daucus carota	Glechoma hederacea s.str
Alchemilla glabra aggr.	Carex ferruginea	Deschampsia cespitosa	Globularia cordifolia
Alchemilla vulgaris aggr.	Carex flacca	Doronicum grandiflorum	Globularia nudicaulis
Alchemilla xanthochlora aggr.	Carex montana	Dryas octopetala	Gymnadenia conopsea
Androsace chamaejasme	Carex nigra	Equisetum palustre	Gypsophila repens
Anemone narcissiflora	Carex ornithopoda	Erigeron uniflorus	Hedysarum hedysaroides
Anthoxanthum odoratum aggr.	Carex pallescens	Euphorbia cyparissias	Helianthemum nummularium s.l.
Anthriscus sylvestris	Carex panicea	Euphrasia hirtella	Helictotrichon pubescens
Anthyllis vulneraria s.l.	Carex sempervirens	Euphrasia minima	Helictotrichon versicolor
Aposeris foetida	Carex sylvatica	Euphrasia salisburgensis	Heracleum sphondylium s.l.
Arabis alpina s.l.	Carlina acaulis subsp caulescens	Festuca ovina aggr.	Hieracium bifidum aggr.
Arnica montana	Carum carvi	Festuca pratensis s.l.	Hieracium lactucella
Arrhenatherum elatius	Centaurea jacea s.str	Festuca quadriflora	Hieracium murorum aggr.
Aster bellidiastrum	Centaurea montana	Festuca rubra aggr.	Hieracium villosum aggr.
Astrantia major	Centaurea scabiosa s.l.	Festuca violacea aggr.	Hippocrepis comosa
Athamanta cretensis	Cerastium arvense s.l.	Fragaria vesca	Holcus lanatus
Bartsia alpina	Cerastium fontanum s.l.	Fraxinus excelsior	Homogyne alpina
Bellis perennis	Cerastium latifolium	Galium album	Hypericum maculatum s.l.
Botrychium lunaria	Chaerophyllum hirsutum aggr.	Galium anisophyllon	Hypochaeris radicata
Brachypodium pinnatum	Cirsium acaule	Galium megalospermum	Juncus effusus
Briza media	Cirsium oleraceum	Galium mollugo	Knautia arvensis

Knautia dipsacifolia s.str
Laserpitium latifolium
Lathyrus pratensis
Leontodon autumnalis
Leontodon helveticus
Leontodon hispidus s.l.
Leucanthemum vulgare aggr.
Ligusticum mutellina
Ligusticum mutellinoides
Linaria alpina s.str
Linum catharticum
Lolium perenne
Lotus corniculatus aggr.
Luzula alpinopilosa
Luzula campestris
Luzula multiflora
Luzula sylvatica
Medicago lupulina
Myosotis alpestris
Myosotis arvensis
Nardus stricta
Parnassia palustris
Pedicularis foliosa
Pedicularis verticillata
Petasites paradoxus
Phleum hirsutum
Phleum pratense
Phleum rhaeticum
Phyteuma orbiculare
Phyteuma spicatum
Picea abies
Pimpinella major
Pimpinella saxifraga aggr.
Plantago alpina
Plantago atrata s.str
Plantago lanceolata
Plantago major s.l.
Plantago media
Poa alpina
Poa cenisia
Poa minor
Poa pratensis
Poa supina
Poa trivialis s.l.
Polygala alpestris
Polygala chamaebuxus
Polygala vulgaris s.l.
Polygonum bistorta
Polygonum viviparum
Potentilla aurea
Potentilla crantzii
Potentilla erecta
Potentilla sterilis
Primula auricula
Primula elatior s.str
Primula veris s.l.
Pritzelago alpina s.str
Prunella grandiflora
Prunella vulgaris
Pulsatilla alpina s.l.
Ranunculus aconitifolius
Ranunculus acris s.l.
Ranunculus alpestris
Ranunculus bulbosus
Ranunculus montanus aggr.
Ranunculus nemorosus aggr.
Ranunculus repens
Rhinanthus alectorolophus
Rhinanthus minor
Rumex acetosa
Rumex alpestris
Rumex crispus
Sagina saginoides
Salix herbacea
Salix reticulata
Salix retusa
Sanguisorba minor s.l.
Saxifraga aizoides
Saxifraga moschata s.l.
Saxifraga oppositifolia
Saxifraga paniculata
Scabiosa lucida
Sedum atratum
Selaginella selaginoides
Senecio doronicum
Sesleria caerulea
Silene acaulis
Silene vulgaris s.l.
Soldanella alpina
Solidago virgaurea s.l.
Stachys officinalis s.l.
Stellaria graminea
Taraxacum alpinum aggr.
Taraxacum officinale aggr.
Thesium alpinum
Thlaspi repens
Thlaspi rotundifolium aggr.
Thymus praecox subsp polytrichus
Thymus pulegioides s.str
Tofieldia calyculata
Tragopogon pratensis s.l.
Trifolium badium
Trifolium medium
Trifolium montanum
Trifolium pratense s.l.
Trifolium repens s.str
Trifolium thalii
Trisetum flavescens
Trollius europaeus
Tussilago farfara
Vaccinium gaultherioides
Vaccinium myrtillus
Vaccinium vitis.idaea
Valeriana montana
Veratrum album subsp lobelianum
Veronica alpina
Veronica aphylla
Veronica arvensis
Veronica chamaedrys
Veronica officinalis
Veronica persica
Veronica serpyllifolia s.l.
Vicia cracca s.str
Vicia sativa s.l.
Vicia sepium
Viola biflora
Viola calcarata
Viola hirta

APPENDIX 2: Environmental variables used in the study

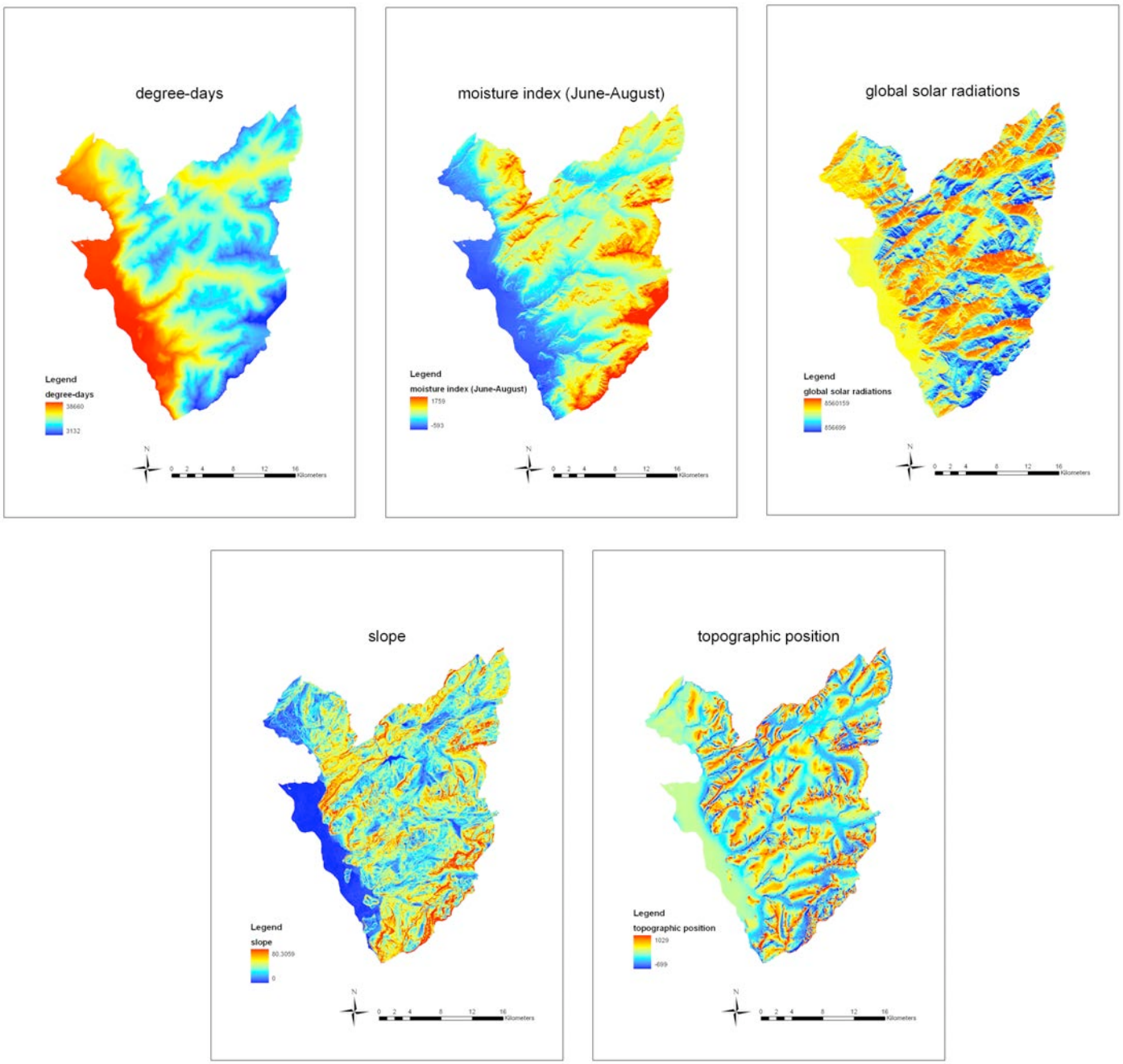

Figure 1A. Maps of the five environmental variables used to model the species richness over the study area (degree-days [°C * day * year^{-1}], moisture index over June to August [mm * day^{-1}], global solar radiations [kJ * m^{-2} * year^{-1}], slope [degrees], topographic position [unitless]). These maps represent the environmental characteristics of the study area.

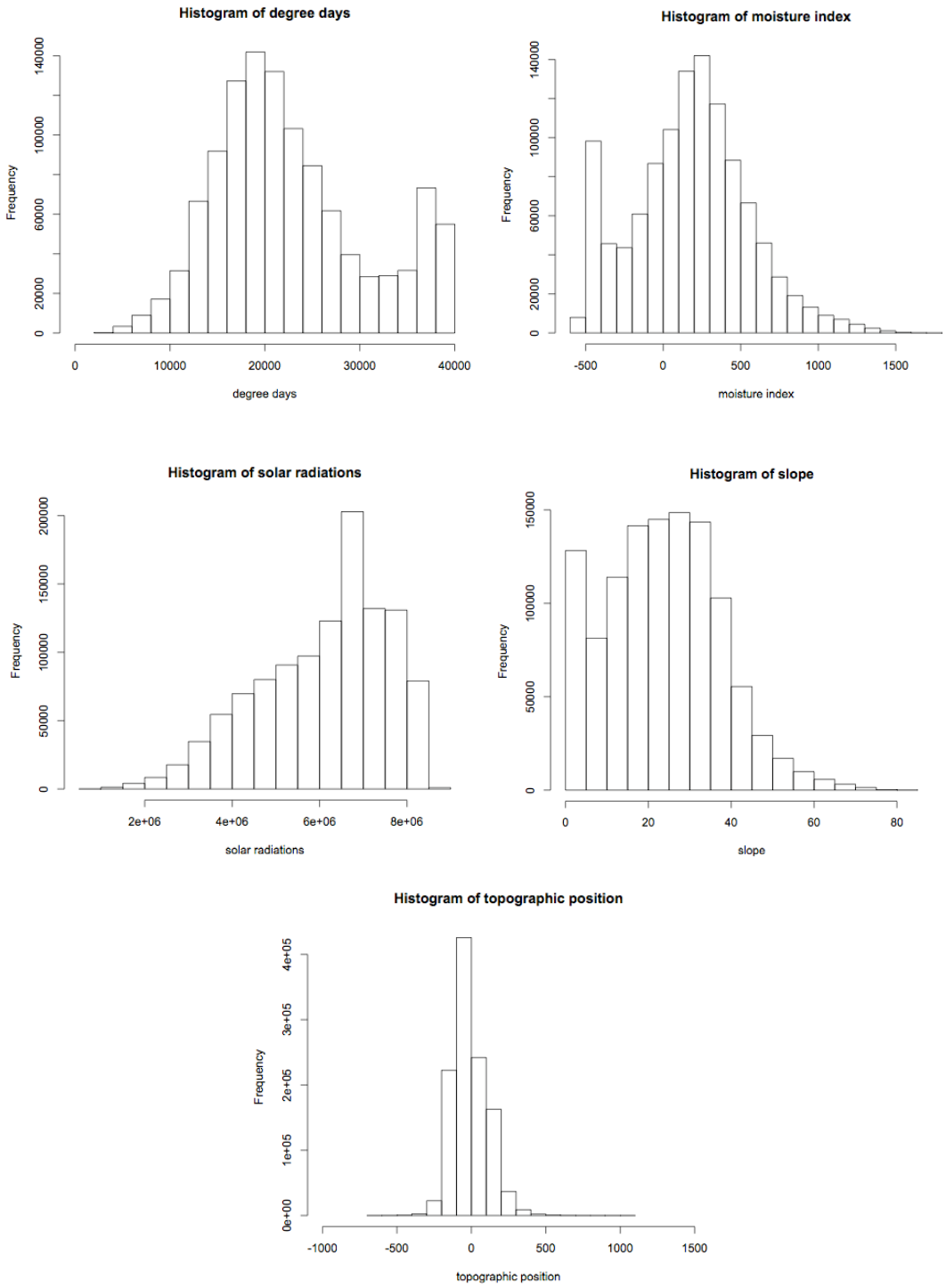

Figure 2A. Histograms representing the distribution of the five variables used in this study.

APPENDIX 3: Correlations between environmental variables

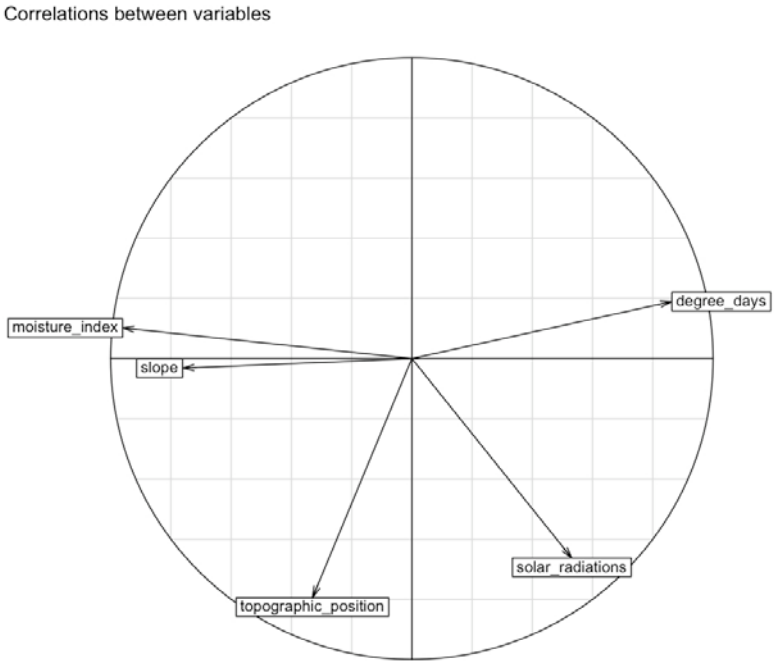

Figure 3A. Graphic representation obtained by a principal component analysis (PCA) of the five predictors used in this study.

Table 2A. Correlation matrix for the five environmental predictors used in this study. Variables are considered as correlated when the correlation is higher than 0.7.

	degree days	moisture index	solar radiations	slope	topographic position
degree days	1	-0.86737	0.18877	-0.51206	-0.27656
moisture index	-0.86737	1	-0.54297	0.20641	0.61577
solar radiations	0.18877	-0.54297	1	-0.28112	0.08596
slope	-0.51206	0.20641	-0.28112	1	0.19615
topographic position	-0.27656	0.61577	0.08596	0.19615	1

APPENDIX 4: Distribution of the number of species observed in the field

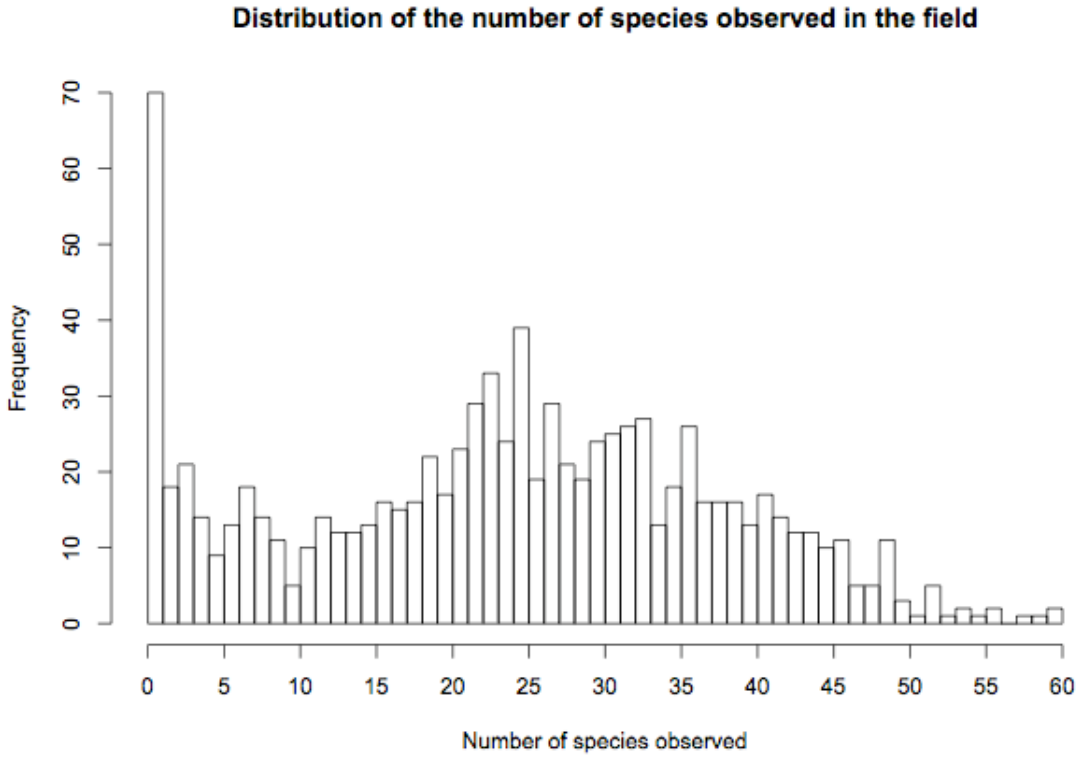

Figure 4A. Histogram representing the distribution of the number of species observed in the field over the 912 sampling plots.

APPENDIX 5: Direct modelling of species richness with Poisson distribution

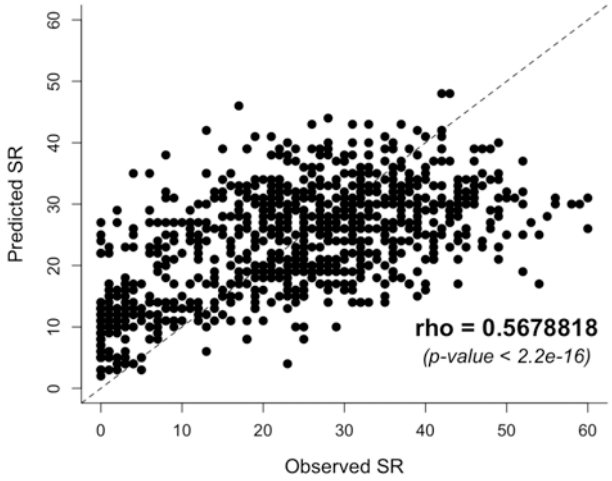

Figure 5A. Scatterplot of the observed versus the predicted SR obtained by the GLM with Poisson distribution. The dotted line represents a pure correlation between observed and predicted SR.

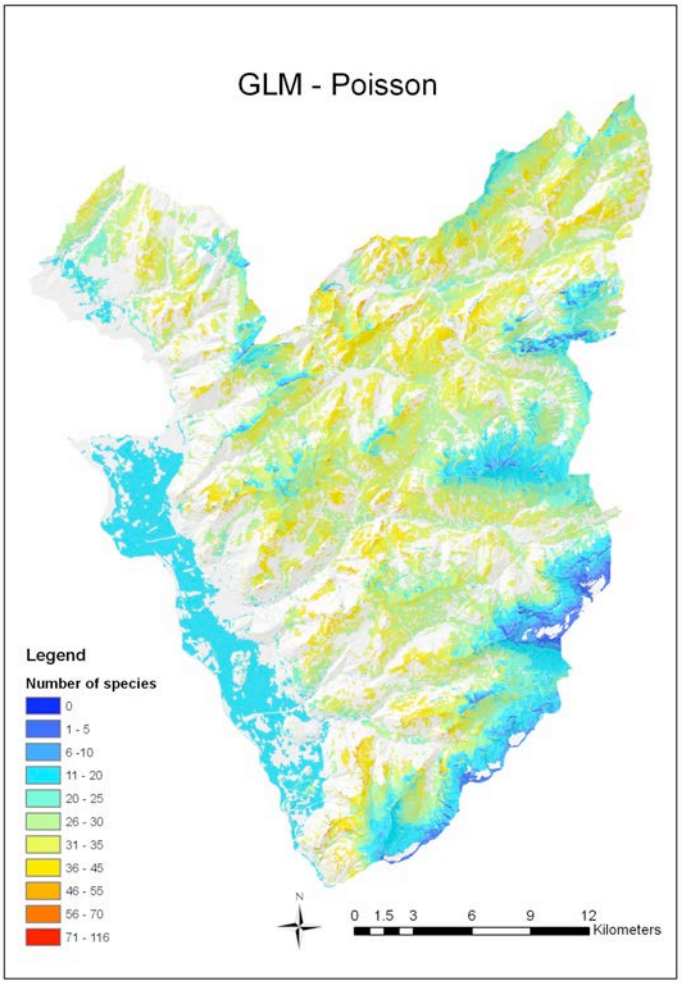

Figure 6A. Spatial distribution of the species richness resulting from the GLM with the Poisson distribution. Regions in white represent areas that were not used for the projections (glaciers, urbanized zones, lakes and forests).

APPENDIX 6: Direct modelling of species richness with zero-inflated negative binomial distribution

Table 3A. Mean measures (mean spearman rank correlation, mean ΔAIC, mean D^2 and mean D^2 adjusted) resulting from the 100-fold cross-validation on the evaluation data set in order to evaluate the GLM and GAM models with the zero-inflated negative binomial distribution.

Models	mean spearman correlation	mean ΔAIC	mean D^2	mean D^2 adjusted
GLM-zero-inflated negative binomial	0.5555092	313.6303	0.0682741	0.06416505
GAM-zero-inflated negative binomial	0.5612684	425.2965	0.08594035	0.08190922

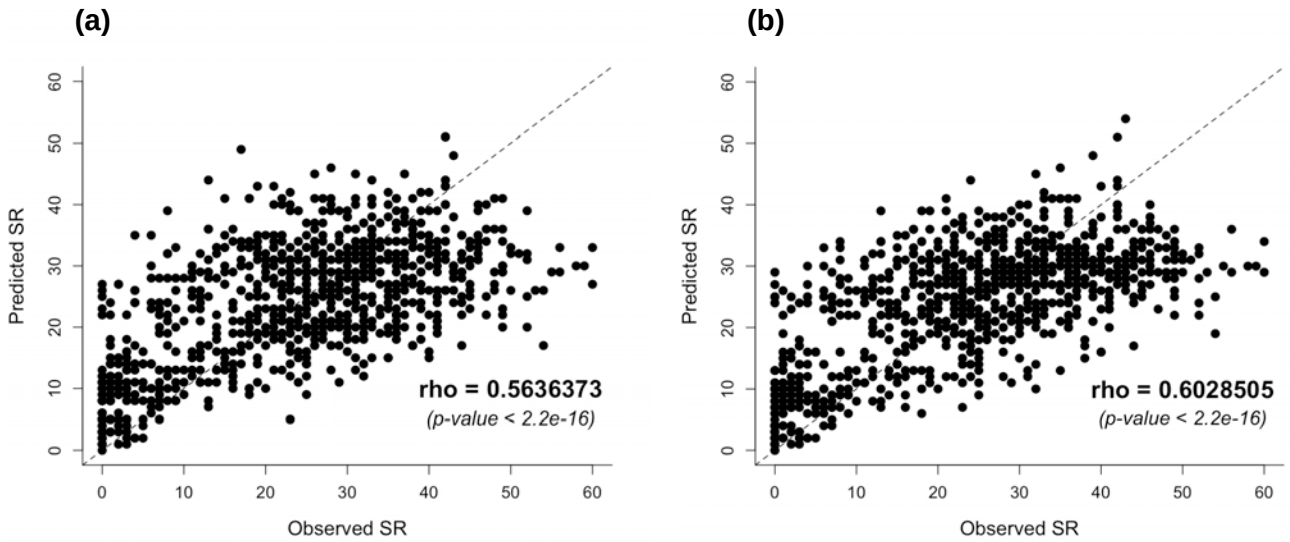

Figure 7A. Scatterplot of the observed versus the predicted SR obtained by (a) the GLM and (b) the GAM with zero-inflated negative binomial distribution. The dotted line represents a correlation of one between observed and predicted SR.

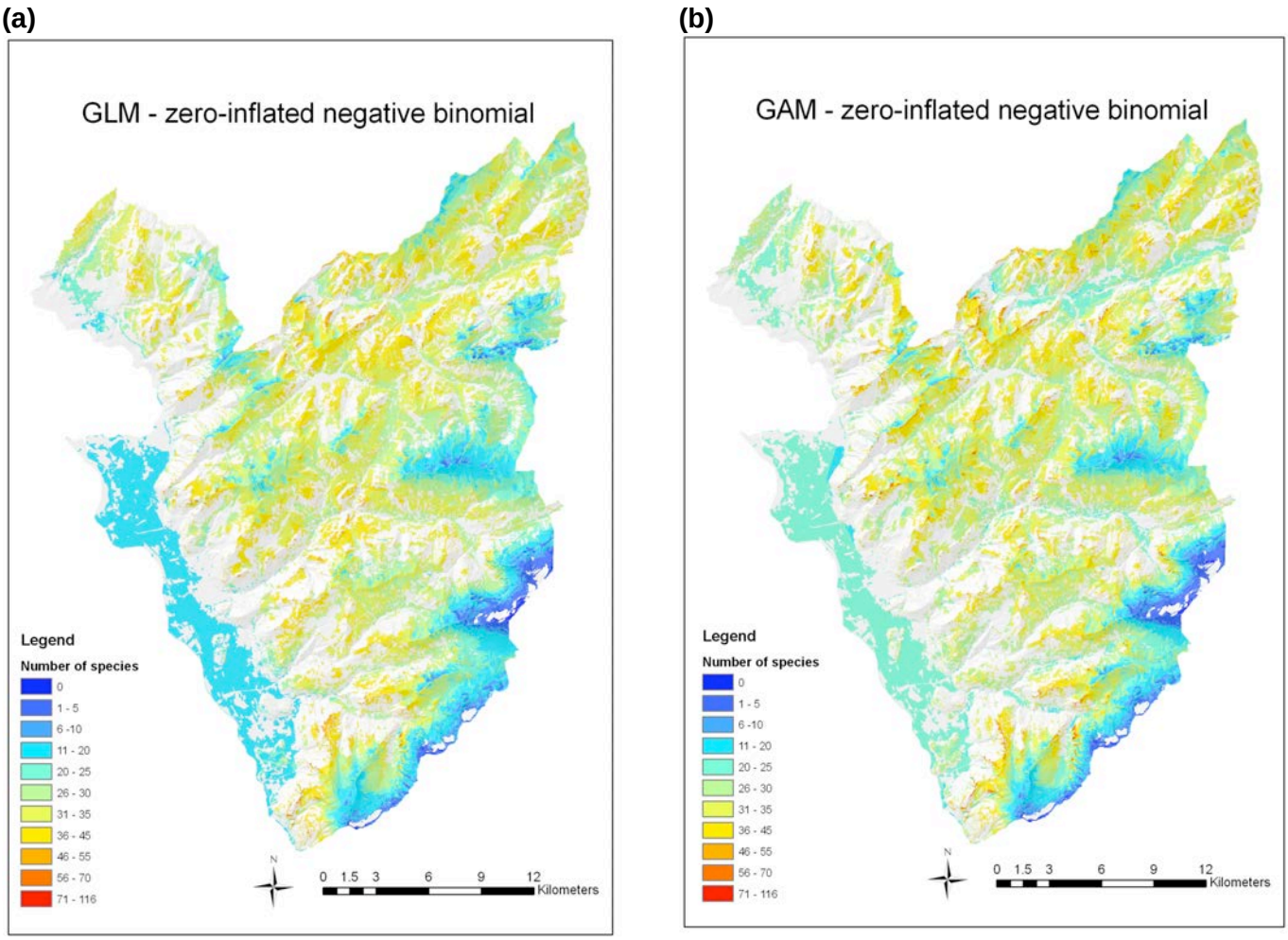

Figure 8A. Spatial distribution of the species richness, resulting from (a) the GLM and (b) the GAM with zero-inflated negative binomial distribution Regions in white represent areas that were not used for the projections (glaciers, urbanized zones, lakes and forests).