UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Ecole de biologie

PREDICTING PLANT DISTRIBUTION: DOES EDAPHIC FACTOR MATTER?

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

Master Thesis of Science in Behaviour, Evolution and Conservation

par

Aline BURI

Directeur : Pr. Antoine Guisan
Superviseur (s) : Pr. Antoine Guisan, Jean-Nicolas Pradervand, Eric Pinto-Figueroa
Expert (s) : Anonymous
Département d'écologie et d'évolution

Janvier 2014

Predicting plant distribution: does edaphic factor matter?

Aline Buri[1*], Jean-Nicolas Pradervand[1], Eric Pinto-Figueroa[1], Thierry Adatte[2], Jorge Spangenberg[2], Erika Yashiro[3], Antoine Guisan [1,4]

1. Department of Ecology & Evolution, Biophore Building, 1015 Lausanne, Switzerland

2. Institute of Earth Sciences, University of Lausanne, Geopolis Building, 1015 Lausanne, Switzerland

3. Department of microbiology, Biophore Building, 1015 Lausanne, Switzerland

4. Institute of Earth Surface Dynamics, University of Lausanne, Geopolis Building, 1015 Lausanne, Switzerland

* Corresponding author: Department of Ecology & Evolution, Biophore Building, 1015 Lausanne, Switzerland; E-mail: aline.buri@unil.ch

Abstract

Aim To investigate the effects of adding edaphic factors in plant species distributions modeling (SDMs) over a complex regional study area.

Location The western Swiss Alps

Methods We used soil parameters collected in the field coupled with topo-climatic data to model two selected edaphic factors: $\delta^{13}C$ and soil pH. By means of an Akaike information criterion (AIC) stepwise selection we derived two predictive models of these edaphic factors for the whole study area. Then we incorporated the modeled edaphic factors to a basic set of topo-climatic predictors (TC). We modeled 260 plant species distributions with three modeling techniques using TC variables alone, TC variables and edaphic factors individually and finally TC variables with both edaphic factors. After that, we assessed the improvement in plant spatial predictions due to addition of edaphic factors.

Results Edaphic factors can be modeled with a $R^2 > 0.80$ over complex area using GLM and GAM statistical techniques. The addition of edaphic factors significantly improved the AUC of the plant models. Adding of soil pH led to better improvement of SDMs' predictive power than adding of $\delta^{13}C$. Nevertheless, the best increase was recorded when both were added as predictors. We also observed a decrease in the over-estimation of species richness calculated by stacked species distribution modeling (S-SDM). This decrease was higher when both edaphic factors were added to SDMs.

Main conclusions Edaphic factors allow an improvement of plant SDMs predictive power. As we showed in this study that they could be generalized in space, we suggest they should be incorporated more frequently when feasible in future studies focusing on plant SDMs.

Keywords species distribution modeling; $\delta^{13}C$, soil pH ; Swiss Alps ; GLM ; GAM

Résumé

But Le but de cette étude est de comprendre l'effet des facteurs édaphiques sur les modèles de distribution d'espèces (SDM).

Localisation Préalpes vaudoises en Suisse

Méthodes Nous avons utilisé des données récoltées sur le terrain ainsi que des données topo-climatiques afin de modéliser deux facteurs édaphiques sélectionnés ; le $\delta^{13}C$ et le pH du sol. A l'aide d'une procédure de sélection stepwise basée sur l'Akaike information criterion (AIC), nous avons dérivé deux modèles pour l'entier de la zone d'étude. Après cela nous avons intégré ces facteurs édaphiques modélisés à un set de prédicteur topo-climatique (TC) basique. Trois différentes techniques de modélisation ont été utilisées afin de modéliser la distribution de 260 espèces de plantes utilisant comme prédicteurs soit seulement les variables TC, soit les variables TC ainsi que les facteurs édaphiques individuellement, soit les variables TC et les deux facteurs édaphiques. L'amélioration du pouvoir prédictif des SDMs dû à l'addition des facteurs édaphiques a été mesuré.

Résultats Les facteurs édaphiques peuvent être modélisés avec un $R^2 > 0.80$ sur des zones complexes en utilisant des techniques statistiques comme le GLM et le GAM. De plus, l'addition de ces facteurs édaphiques modélisés améliore l'AUC des modèles de distribution de plantes. L'addition du pH du sol mène à une amélioration plus haute de l'AUC que l'addition de $\delta^{13}C$. Néanmoins, la plus grande amélioration est le résultat de l'addition des deux facteurs édaphiques simultanément. Un autre résultat montre que la surestimation de la richesse spécifique calculée par stacked species distribution modeling (S-SDM) est diminuée par l'addition de facteurs édaphiques. Cette diminution est augmentée lorsque les deux facteurs édaphiques sont ajoutés aux SDMs

Principales conclusions Les facteurs édaphiques permettent une amélioration du pouvoir prédictif des SDMs. Etant modélisable sur des zones complexes, nous suggérons que ces variables soient prises en compte, dans la mesure du possible, lors de la création de modèle de distribution d'espèces de plantes dans de futures études.

Introduction

The principle of holocoenotic environment (Billings, 1938) considers the environment-plant system as a dynamic unit which varies as a whole. Therefore, plants are in interaction between themselves and with their environment, but the environmental factors are also interacting between themselves. Plant growth and plant physiological processes are limited by a multitude of environmental factors (Cain, 1944). Thus, the ensemble of habitats suitable for a plant species is defined by the niche principle as a hypervolume of environmental conditions in which positive growth rate can be maintained (Hutchinson, 1957). Accordingly, it suggests that we have to take into account simultaneously many environmental factors to understand plants' distribution. Good (1964) proposed a classification of the importance of these environmental factors in shaping plant distributions; climatic ones are the most important followed by edaphic ones and, at the end, biotic interactions.

It has been confirmed since then that physiological tolerance to climatic factors is the main driver of plants distributions (Woodward, 1987) and that climatic variables are therefore primary factors for predicting plant distributions (Box et al., 1993; Shao & Halpin, 1995; Heegaard, 2002). These climatic layers are easily obtained by interpolation of meteorological station measurements using Geographic information system (GIS) (Ninyerola et al., 2000). Thus, they are widely used to predict plant species distribution through spatial distribution modeling (SDM). (Thuiller et al., 2005; Hijmans & Graham, 2006)

Edaphic factors were ranked secondly important by Good (1964) and their importance is well known for plants: they constitute the belowground environment. Some edaphic factors directly influence plants physiological processes, such as ions (*e.g* phosphorus, calcium, nitrogen) used as nutrients (Aerts & Chapin, 1999). Other factors can determine water and nutrients supply, such as clay and organic matter content (Beadle, 1953; Gobat et al., 2004). For example many ions' availability mostly depends on soil acidity (Gobat et al., 2004). Soil pH also directly affects plant growth (Austin, 1980). Moreover, plants can be physically affected by edaphic factors; soil rockiness and particle size distribution act on plant by supporting them, but also by limiting their roots growth (Martre et al., 2002). However, plants also reciprocally influence some edaphic factors. $\delta^{13}C$ represents the ratio of stable isotopes ^{13}C:^{12}C (Battle et al., 2000; Keeling et al., 2011). Measuring this ratio in the soil allows the

identification of the plants community $\delta^{13}C$ (Andreux et al., 1990). Plants can discriminate between ^{13}C and ^{12}C during the photosynthesis, thus they can act on this soil $\delta^{13}C$.
So far, only few studies have focused on SDMs including both climatic and edaphic variables, despite previous calls and evidences (*e.g.* Pinto & Gégout, 2005; Coudun et al., 2006; Austin & Van Niel, 2011) showing that edaphic factors are needed as predictors to improve models of species distributions (in the examples cited, for tree species).

Several reasons can explain why edaphic factors are rarely included in SDMs. Measuring these edaphic factors is usually time consuming and costly. Furthermore, to allow a wide use of edaphic factors in SDMs, layers with soil proprieties must be generalized over the geographic extent of the study area. Different techniques can be applied to generalize soil proprieties over the whole area. Yet, most of these rely on spatial interpolations, which assume some sort of diffuse processes from emission sources behind the observed patterns. However, this is unlikely to apply to soil parameters in rugged landscape, like mountains. As an alternative, a few authors tested the use of predictive modeling approaches, using statistical techniques like generalized linear models, classification and regression trees, neural networks, or fuzzy systems that differ greatly from geostatistical interpolations (Park & Vlek, 2002; McBratney et al., 2003; Bishop & McBratney, 2006). The paucity of predictive approaches to soil mapping has prevented yet the development of a general method to model soil properties in geographic space. As a result, most of the existing studies have not taken topographic parameters into account, although these would represent relevant dimensions in many study areas for the creation of soil maps.

The main objective of this study was thus to investigate the effects of adding edaphic factors in topo-climatic models of plant distributions (SDMs) over a regional study area encompassing a complex topography, using empirical data on plant distribution and soil parameters collected in the field. Using a three-step method, we first selected edaphic factors with possible effects on plant distributions. Then, we modeled the geographic variation in the values of two among the most important edaphic factors. We finally tested the effect of adding these edaphic factors as predictors in plant SDMs and assess the improvement in spatial predictions across the study area. We expected that:

1: Variations in edaphic factors can be modeled over large and complex study areas using predictive modeling.

2: Adding edaphic factors would improve the predictive power of plant SDMs in the study area and increase the accuracy of species richness predictions derived from single SDMs.

Methods

Study area and data collection

The study area (46°10’-46°30’N ; 6°50’-7°10’ E, Fig. 1) is located in the Western Swiss Alps covering ca. 700 km^2. The elevation gradient ranges from 375 m to 3210 m asl. The climate is temperate with annual precipitation varying from 1200 mm at low elevation to 2600 mm at high elevation, whereas annual temperature ranges from 8°C at low elevation to -5 °C at high elevation (Bouët, 1972). The bedrock is mainly calcareous but flysch, gypsum, granite or marlstone rock can be found. Different soil types are present; they range from deep brown earth (Brunisols) to leached brown earth but also more fine and thin soil like the lithosols (Gobat et al., 2004; Baize et al., 2009). The vegetation belts’ succession along the altitudinal gradient is typical from the calcareous Alps, from lowest to highest elevations: a colline belt with deciduous forests, a montane belt with mixed forests, a subalpine belt with coniferous forests, an alpine belt with heath, meadow and grassland vegetation, and finally a nival belt with sparse vegetation of high-elevation species (Randin et al., 2006; Aeschimann & Burdet, 2008). The anthropic influence on soils and vegetation is mainly due to agriculture in the form of pasturing, mowing, manuring and occurs from lowland to subalpine belt, decreasing gradually along the altitudinal gradient.

The sampling sites were selected in non-forest zone of the study area according to a random-stratified balanced strategy (Hirzel & Guisan, 2002). The sampling strategy was based on previous vegetation inventories done between 2002 and 2009 on 912 plots. The random-stratified strategy used to select the vegetation plots was based on elevation, slope and aspect as stratifying variables. Plants species absence-presence were recorded in 4 m^2 square (Randin et al., 2006; Dubuis et al., 2013). Soil information was collected during summer 2012 and 2013 among 242 of the 912 vegetation plots (*i.e* 102 plots in 2012 and 140 plots in 2013) keeping the same sampling design. For each plot, a 2 m^2 square was delimited. Soil cores were extracted from topsoil (5 cm deep) in the 4 corners and the center of the square, and then homogenized in a sterile plastic bag. A subsample of soil was put in ashed aluminum foil and frozen with liquid nitrogen, the remains was let in the plastic bag and stored on ice. Soil temperature was recorded using a laser thermometer inside the 5 dug holes. Samples were carried in the lab right after collection. Chemical analyses of the soil samples stocked on ice were directly proceeded whereas the frozen soils samples were stocked at -80 °C. They were chemically analyzed after field season.

Analyzing the soil chemical composition

We proceeded to different analyses in order to characterize the soil's chemical composition. The following edaphic factors were measured: soil pH, electro-conductivity (EC), soil water content, phosphorus (P), mineral carbon (MINC), total organic carbon (TOC), oxygen index (OI), hydrogen index (HI), the temperature at maximum release of hydrocarbons (Tmax), carbon (C), hydrogen (H), nitrogen (N), carbon and nitrogen isotopic ratio (δ13C and δ15N), major elements (SiO_2, TiO_2, Al_2O, Fe_2O, MnO, MgO, CaO, Na_2, K_2O, P_2O, Cr_2O_3, NiO), mineralogy (Phyllosilicate, Quartz, Feldspath-K, Plagioclase-Na, Calcite, Dolomite, Goethite, Ankerite, Indoses) and organic mater (OM).

We first used the non-frozen soil samples to measure the soil pH in a 1:2.5 soil/water suspension with the pH meter method (Page, 1982). We also conducted an electrical conductivity of soil analysis using 1:1 and 1:5 aqueous extractions. Soil water content was calculated following USDA-NRCS protocols (Soil Survey Staff, 2009).

The subsets of soil samples stored in aluminum foil at -80°C were initially freeze-dried on LYOVAC machine. Dry samples were then sterility transferred in glass jars and stored at -20°C. To prepare samples for analysis, we sieved at 2 mm and ground the soil into powder. A subset of each sample was also decarbonized with 10% HCl. Total phosphorus content was determined by colorimetric analysis after mineralization at 550°C with $Mg(NO_3)_2$. The characterization of the bulk organic matter, including MINC, TOC, OI, HI and Tmax, was done using a Rock-Eval 6 machine with a standard whole-rock pyrolysis method (Lafargue et al., 1998). Nitrogen, carbon and hydrogen contents analysis were performed using a Carlo Erba CNS2500 CHN Elemental Analyzer coupled with a Fisons Optima mass spectrometer (Tamburini et al., 2003). The carbon and nitrogen isotopic compositions were established by flash combustion of the acidified samples on a Carlo Erba 1108 (Milan, Italy) elemental analyzer (EA) associated with a Thermoquest/Finnigan MAT Delta S (Bremen, Germany) isotope ratio mass spectrometer (IRMS) (Spangenberg et al., 2006). The major elements' composition was defined using a X ray fluorescence analysis and we used an X ray diffraction analysis to determine soil mineralogy. Finally, we performed a loss of ignition (LOI) at 1050°C to determine the organic matter's contents. For more details about chemical analyses and edaphic factors, see Appendix S1 in supporting information.

Modeling species distributions using edaphic factors

To test for the importance of edaphic factors in species distribution models, we used a three-step method. First we analyzed the importance of edaphic factors in plant species distributions using the soil data collected in 2012 on 102 sites. In order to work on the whole set of sites and thus the whole set of plant species we modeled two important edaphic factors in plant species distribution models over the whole study area. After that, we extracted the information of our soil models for the remaining 810 sites where soil data was not collected. Finally, we tested the effect of adding edaphic factors in plant SDMs on the whole study area.

Selecting edaphic factors to models

To test the relative importance of the edaphic factors, we performed plant SDMs using all of the 40 edaphic factors as predictors (Guisan & Zimmermann, 2000). We modeled the distribution of the 42 plant species having more than 20 occurrences in the 102 sites visited in 2012 using the biomod2 library (Thuiller et al., 2013) of the R software (version 3.0.2). Three statistical modeling techniques were used: generalized linear models (GLM, McCullagh & Nelder, 1989), random forest (RF, Breiman, 2001) and generalized additive models (GAM, Hastie & Tibshirani, 1990). An Akaike information criterion (AIC) stepwise GLM and GAM with degree of freedom up to 4 and splines smoothers were calibrated. All four models used a logistic link function. Ensemble modeling was then performed to limit the variability due to the different modeling techniques (Araújo & New, 2007). We recorded the mean variables' importance of each edaphic factor among models and ranked them by the median in an increasing order through a boxplot. We used these data to compare the influence of the different edaphic factors on plant SDMs.

We selected $\delta^{13}C$, the edaphic factors having the highest importance among the 40 factors tested. $\delta^{13}C$ being unusual in ecology field, some explanations are necessary to understand its relation with plants. $\delta^{13}C$ measures the ratio of stable isotopes ^{13}C :^{12}C in soil (Battle et al., 2000; Keeling et al., 2011). Atmospheric carbon dioxide contains two types of stable carbon isotope: ^{12}C and ^{13}C. Plants can discriminate between these two types of stable isotopic carbon during the photosynthesis (Smith & Epstein, 1971). Plants incorporate more ^{12}C than ^{13}C in their biomass, and therefore, their $\delta^{13}C$ is lower than the atmospheric one. Nevertheless, the rate of discrimination between ^{13}C and ^{12}C depends of the photosynthetic pathway type (Farquhar et al., 1989). Hence, plants $\delta^{13}C$ vary according to their photosynthesis pathway type (CAM, C_3, C_4). It has been shown that $\delta^{13}C$ also vary between

C_3 plants due to their differential proportion of stomatal and mesophyll resistance to $^{13}CO_2$ uptake (Farquhar et al., 1980; O'Leary & Osmond, 1980; Vogel, 1980; Evans et al., 1986). $\delta^{13}C$ from the upper soil profile is similar to that of the plant community, thus it reflects the relative contribution of the different plant species photosynthetic pathways to the primary productivity of the plant community (Andreux et al., 1990; Balesdent et al., 1996; Boutton & Yamasaki, 1996; Gillson et al., 2004).

As the variables following $\delta^{13}C$ were ranked very close in the terms of relative importance in the global model, we decided to select, for building soil map, the one known to be the most proximal for plant distribution (Roem & Berendse, 2000; Pärtel, 2002; Chytr et al., 2003). Soil pH represents the ionic concentration of the soil (Lozet & Mathieu, 1997). It is known as a resource variable for plants because of its controls on the mineral uptake of plants (Duchaufour, 1989; Hossner, 2008). It is also known as a direct variable because of its physiological effect on the plant growing (Austin, 1980). Soil pH also constraints microorganism and soil wildlife activity, soil heavy metal concentration and humus formation (Bolan & Kandaswamy, 2005). Moreover, it has been shown that soil pH influences plants SDMs (Dubuis et al., 2013). This previous study has been carried out on a small data set, thus using soil pH in our large-scale study would allow confirming their finding. Consequently, soil pH is a good comparison point to test the effect of other edaphic factor on SDMs.

Modeling edaphic factors for the whole study area

To model the two edaphic factors, we used stepwise GLMs and GAMs. The stepwise selection selects the most significant predictors among a large set of potential predictors (Table 1) in the models by sequentially adding or removing variables in the model. The selection of the variables was based on Akaike's information criterion (AIC, Akaike, 1974). We took care not to use correlated predictors. Adjusted R^2 (Nagelkerke, 1991) was calculated for each model to estimate the model fit. A repeated cross-validation procedure was performed using 80% of the data for calibration and 20% to test the prediction. Pearson correlation between predicted and observed value was calculated at each repetition (run). A mean of this correlation was done after 1000 runs. Model with the highest adjusted R^2 and correlation from repeated cross-validation were selected to continue the analysis. We calculated the relative influence of each predictor (x_i) using the formula:

$$\Delta R^2_{xi} = R^2_{x} - R^2_{-xi}$$

Where R^2_x is the R^2 of the full selected model, R^2_{-xi} is the R^2 of the selected model minus the variable x_i.

Outputs of the two fitted models were at a resolution of 25 m x 25 m for the whole study area.

Testing the importance of edaphic factors in species distribution modeling of the whole study area

Using the biomod2 R package (Thuiller et al. 2009), we modeled the species distributions of the 260 plant species having more than 20 occurrences over the 912 plots covering the whole study area. In total, we ran four sets of models using different set of predictors. We used the same parameters and modeling techniques as in the previous section. The first set of models used four basic topo-climatic (4 TC) predictors known to be important for plant distribution in the study area (Engler et al., 2009; Randin et al., 2009; Pellissier et al., 2010): the summer air temperature, the summer radiation, the slope and the topographic position (See Appendix S2 for more information). This step gave us the model quality with common topo-climatic predictors. Then we ran two sets of models with the 4TC variables and the two edaphic factors used individually (TC + edaphic factor 1, TC + edaphic factor 2). Finally, we ran a last set with the 4TC variables and both edaphic factors. To test if the difference between models having 4, 5 or 6 predictors was not simply due to the difference between the numbers of predictors, we also ran the TC models but with randomized soil predictors (TC + randomized edaphic factor 1, TC + randomized edaphic factor 1 + edaphic factor 2).

For each set of models we recorded the variables importance of used predictors given by biomod2. We also calculated the area under the curve (AUC, Fielding & Bell, 1997) of a receiver operating characteristic (ROC) plot for each model. The number of species predicted to occur on each plot was estimated by stacking individual species distribution models (S-SDM) (Ferrier & Guisan, 2006; Guisan & Rahbek, 2011), thus predicting patterns of species richness. We then compared the AUC per plot between the different sets of models. We also compared the predicted species richness per plot with the observed species richness. Finally, we calculated the percentage of average over-estimation in species richness predictions of the three different sets of models.

Results

Analyzing the soil's chemical composition

Average values for soil pH and $\delta^{13}C$ were 6.44 and -27.10 ‰V-PDB. Soil temperatures ranged from 5.60°C to 24°C. Mean values for $\delta^{15}N$, H and soil dry weight were

1.22 ‰ AIR, 1.26% and 159.81 g. MnO percentages were ranging from 0.02 wt-% to 0.34 wt-% (Table 2).

Modeling species distributions using edaphic factors

Selecting edaphic variables to models

Among the 40 variables used in the models, the ratio of stable isotopes ^{13}C :^{12}C ($\delta^{13}C$) showed the greatest impact on plant SDMs (Fig. 2), followed by soil temperature and $\delta^{15}N$. Soil pH occurred at the 7th position in the variable importance ranking according to their median (Fig. 1). Variables taking place at the second to seventh position (*i.e* soil temperature, $\delta^{15}N$, H, dry weight, MnO, soil pH) are ranked closely. We selected $\delta^{13}C$ and pH as explanatory variables to model over the whole study area.

Modeling edaphic factors across the whole study area

$\delta^{13}C$ model

$\delta^{13}C$ model with the highest fit is a GLM with a Gaussian distribution where each predictor was allowed up to second order. The model was calibrated using 102 plots information. It kept five variables: winter temperature (WT), autumn moisture index (AMind), spring radiation (Srad), normalized difference snow index (NDSI) and the elevation (dem) as predictors. (Formula: $\delta\ ^{13}C = -\ 30.06 - (4.86e^{-02}*WT) - (3.86e^{-03}*WT^2) - (4.44e^{-03}*AMind) + (2.63e^{-06}*AMind^2) + (2.38e^{-}04*Srad) - (6.08e^{-09}*Srad^2) + (1.16*NDSI) + (7.05e^{-01}*NDSI^2) + (2.90e^{-03}*dem) - (5.29e^{-07}*dem^2)$, adjusted $R^2 = 0.83$, cross validation correlation = 0.86). Elevation and autumn moisture index were the most influent factors among the five predictors ($\Delta R^2 = 0.016$ and $\Delta R^2 = 0.0148$, Appendix S3). $\delta^{13}C$ was then predicted for the whole study area using the fitted GLM (Fig. 3). Predicted values ranged from -31.02‰ to -22.78‰, which is coherent with the literature saying that $\delta^{13}C$ in temperate soil range from -34‰ to -22‰ due to plants using C_3 pathway to fix CO_2 during photosynthesis (Vogel, 1993).

Soil pH model

The fitted model for soil pH was a GAM model, with Gaussian distribution and a smoothing splines function allowing 4 degrees of freedom using 6 variables: winter moisture index (WMind), spring radiation (Srad), normalized difference vegetation index (NDVI), normalized difference snow index (NDSI), site water balance (SWB) and topographic position (topo). (Formula: soil pH $= 10.48 - (1.66e^{-05}*Srad\ ^4) - (8.772916e^{-04}*WMind^4) - (1.72*NDVI^4) - (6.82e^{-04}*SWB^4) + (2.19*NDSI^4) - (3e^{-03}*topo^4)$, $R^2 = 0.82$, cross-validation correlation = 0.80). It was calibrated using 242 plots information. Topographic position ($\Delta R^2 =$

0.039, Appendix S3) was the most influent factor in shaping soil pH distribution followed by winter moisture index (ΔR^2 = 0.036, Appendix S2). The predicted values of soil pH for the whole study area using the fitted model ranged from 3.51-12.66 with the majority of values ranging from 5 to 8 (Fig. 4). It was consistent with the pedological literature regarding calcareous bedrock saying that soil pH usually ranges from 4 to 7,5 (Gobat et al., 2004).

Testing SDMs improvement

Summer air temperature was the most important predictor in the topo-climatic SDMs, followed by the slope, the summer radiation and the topographic position (Fig. 5 a). In the SDMs using the four topo-climatic variables and $\delta^{13}C$ as predictors, summer air temperature was still the most important predictor. It was directly followed by the edaphic factor, then by the slope, the summer radiation and the topographic position (Fig. 5 b). The same pattern appeared when the soil pH was added to the basic set of the 4 TC predictors (Fig. 5 c). When both edaphic predictors were added to the basic set of TC predictors, summer air temperature was still the most important predictor and edaphic predictors followed; $\delta^{13}C$ in second position and soil pH in third position followed by the slope, the summer radiation and the topographic position (Fig. 5 d).

We found that SDMs having one edaphic variable randomized (TC + randomized pH, TC + $\delta^{13}C$ + randomized pH) had an AUC value either equal or smaller than the same SDMs without this randomized variables (TC, TC + $\delta^{13}C$). The SDMs based on the four topo-climatic (TC) variables were improved by adding edaphic factors to the set of predictors. Adding of soil pH led to a better AUC improvement than adding of $\delta^{13}C$. Nevertheless, the best AUC increase was recorded when both were added as predictors (Fig. 6). More precisely, adding δ13C to the basic set of TC predictors increased the AUC of SDMs for 72% of the 260 plant species. The average AUC increase was 0.012 representing 1.36% of increase; significant increase was observed for RF and GAM modeling techniques (Wilcoxon-test, respectively p-value <0.001 and 0.023, Table 3). SDMs based on 4TC and soil pH showed an increase in AUC compared to basic SDMs for 81% of plant species. As with $\delta^{13}C$, a significant increase in AUC was observed for RF and GAM modeling techniques (Wilcoxon-test, respectively p-values <0.001, Table 3). Here, the average AUC increase was 0.016, it represents a 1.9% increase. Finally, SDMs with both edaphic factors as predictors showed an average AUC increase of 0.024 representing a 2.88% increase. 88% of plant species showed an AUC increase. The three modeling techniques were significantly improved (Table 3).

Comparing observed with predicted species richness showed that models over-estimated species richness. SDMs with 4 TC variables as predictors had an average over-estimation of 150%. Adding $\delta^{13}C$ to the set of predictors decreased the over-estimation by 7%, whereas adding soil pH decreased it by 14% and adding both decreased it by 18% (Table 4). The GAM SDMs with $\delta^{13}C$ and soil pH as predictors showed the higher average reduction with a 35% decrease in AUC (Fig.7).

Discussion

In this study we showed that $\delta^{13}C$ and soil pH could be modeled over large and complex area with respectively R^2 = 0.83 and R^2 = 0.82 using GLM and GAM statistical techniques. Moreover, when comparing plant SDMs using or not the modeled edaphic factors as predictors, we showed that edaphic factors could significantly improve 88% of the plant SDMs. When both edaphic factors were added as predictors we observed an average AUC increase of 2.88%. We also observed an 18% average decrease in the over-estimation of species richness calculated by S-SDMs when both edaphic factors were added to SDMs.

Modeling edaphic factors

One of the objectives of this study was to obtain high-resolution edaphic maps, which could be used as environmental variables to predict the distribution of plant species. Few studies focused on $\delta^{13}C$. However, Suits et al., (2005) predicted the global plant $\delta^{13}C$ at a 1° x 1° spatial resolution (corresponding to ca. 111 km x 111 km at the Equator) using a revised simple biosphere model (SiB2) coupled with a carbon isotope discrimination model. They used as input assimilated meteorology data (*i.e* wind speed, temperature, relative humidity, solar radiation, precipitation) as well as satellite-derived vegetation characteristics, canopy height, soil properties and distribution of C3 and C4 plants. Our study focused on a much finer resolution modeling (*i.e.* 25 m x 25 m) in order to incorporate the predicted edaphic values into fine scale SDMs. Still, the GLM fitted to our data allowed a fairly accurate prediction of this factor across our study area. Furthermore, the predictors selected by the stepwise GLM belong to the same categories than those used by Suits et al., (2005) (*i.e* temperature, moisture, radiation) except for the elevation and NDSI predictors. However, previous studies showed that isotopic $\delta^{13}C$ is influenced by the plant photosynthetic pathway (Smith and Epstein, 1971), irradiance (Ehleringer et al., 1986), water availability (Stewart et al., 1995), air temperature (Ekblad et al., 2005) but also elevation (Körner et al., 1991). However, obtained maps slightly differ. Suits et al. (2005) found predicted $\delta^{13}C$ ranging from

-27‰ to -25‰ for the Swiss area. We found predicted values ranging from -31.02 to -22.78‰. This larger range could be due to the higher precision of our study leading to a better capture of the $\delta^{13}C$ variation across the rugged landscape investigated. It may also result from a too high flexibility of our model. Indeed, observed values in our study area were ranging from -29.50 to -24.37‰. Nevertheless, it is known that $\delta^{13}C$ in temperate soil range from -34 to -22‰ due to plants using C_3 pathway to fix CO_2 during photosynthesis (Vogel, 1993). Our predicted values fit with this result. In particular our isotopic $\delta^{13}C$ map showed higher values at high elevation, which is consistent with the global observation done by Körner et al., (1988).

Regarding the soil pH, Hengl et al., (2004) attempt to predict pH in a Croat study area of 2500 km^2 (n=135 plots). They used kriging and regression-kriging (*i.e.* combining either simple or multiple-linear regression model with ordinary or simple kriging of the regression residuals), but they failed to have good prediction. Castrignanò et al., (2011) used a DEM to improve pH prediction calculated by several kriging methods. They obtained satisfactory to good models in a 15 km^2 alpine doline-type landscape made of carbonatic bedrock located at 1900 m asl. on the Italian Alps. They predicted soil pH and obtained values ranging from 3.85 to 7,5 over the five models. Our predicted values are ranging from 5 to 8. Both are in accordance with the pedological theory saying that soil pH varies from 4 to 7,5 in temperate regions (Gobat et al., 2004). Nevertheless, our study area tends to have a more basic soil pH compared to an average expectation. It is likely due to the fact that western Swiss Alps have primarily calcareous bedrock, which usually ranges from basic (8.2) to neutral (7) soil pH (Gobat et al., 2004; Soil Survey Staff, 2009). At high elevation where soil is shallow, calcareous bedrock may also influence more the soil pH. Indeed, soil contains absorbent complexes that capture ions, usually H^+. The more H^+ are fixed, the higher the pH is. When bedrock is calcareous, absorbent complexes near to bedrock are full of Ca^+ instead of H^+. Thus, soil H^+ concentration is lower near bedrock than far from it. Hence, in shallow soil, this basic influence of calcareous bedrock may be more visible than in wide soil. We can see on our pH map that values are high at really high elevation.

It is to notice that Castrignanò et al., (2011) investigated a small area with a reasonable number of plots (15 km^2, 110 plots). We are working with an elevation gradient ranging from 300 m to 3000 m asl in a 700 km^2 complex landscape with a smaller density of points (although we have in absolute more plots). High quality interpolations (*e.g.* kriging) of mountainous areas are difficult because huge variations exist over small distances that cannot be captured by geographic positions. Thus, for interpolation methods in such landscape, a very high number of plots would be needed to capture all the terrain variation. Here, if we wanted to keep the same plots number per m^2 than Castrignanò et al, (2011) we should have sampled 5’133 plots. Therefore, using modeling techniques appeared more appropriate in our situation. A single other author tested the use of modeling techniques to predict pH, although on a smaller area. Park & Vlek, (2002) used artificial neural network (ANN), regression tree (RT) and generalized linear model (GLM) approaches to predict soil components including pH in three dimensions. These authors used soil type (8 classes), vegetation occurrences (11 species) and terrain attributes (elevation, slope gradient, aspect, plan curvature, profile curvature, upslope area, curvature and wetness index) as predictors. Their 30 km^2 study area was in the UK (n=54 plots, but with 10 different depth samples each time). They also obtained the best pH model with the generalized regression approach (GLM). In our model the stepwise process selected the same types of predictors. However, in our case, radiation and moisture were additionally selected. Still, our results reflect overall previous findings and confirm the feasibility to map soil properties in space through predictive modeling approaches instead of interpolations in complex landscapes.

Improvement of plant SDMs

Only a few studies so far successfully incorporated edaphic factors in plant SDMs (Pinto & Gégout, 2005; Coudun et al., 2006; Coudun & Gégout, 2007) focusing mainly on trees or shrub species (e.g. *Acer pseudoplatanus, Acer campestre, Carpinus betulus, Fraxinus excelsior, Pinus sylvestris* and *Vaccinium myrtillus*). In the present study, we similarly expected to improve SDMs for non-woody vegetation, because soil is an important part of the environment for all plants. Here we also tested how adding soil variables improved species richness predictions. SDM and richness predictions’ improvement varied with soil variables. Our results suggest that edaphic factor do matter in non-woody plant SDMs.

In accordance with previous studies (Box et al., 1993; Heegaard, 2002), the importance of the variables in our models showed a constant pattern. The importance of the edaphic factors in the models was lower than temperature, always in first position, but still greater than the topographic predictors corroborating previous results showing that climate plays the major role in plant SDMs (Shao & Halpin, 1995; Pradervand et al., 2013). Nevertheless, edaphic factors were always at the second position suggesting a better explanatory power from those variables than the topologic variables. This result is supported by other studies: Coudun & Gégout, (2007) used a GLM forward stepwise procedure to select best predictor for *Vaccinium myrtillus* among 7 potential predictors including 5 TC and two edaphic variables. The predictors finally selected were mean annual temperature, measured soil pH and measured C:N ratio. Coudun et al., (2006) used a similar stepwise GLM to model the distribution of *Acer campestre*. Measured pH was the first variables selected, followed by Autumn precipitation and actual Thornthwaite's evapotranspiration.

Not only edaphic factors show high importance among predictors used in SMDs, but they also lead to an improvement of plant SDMs. Similarly, Coudun et al., (2006) found an AUC increase of 0.17. In this study, the increase of AUC due to $\delta^{13}C$ was lower than the one due to soil pH. We can explain this difference by the fact that pH influences vegetation by acting both as a resource and as a regulator variable, while $\delta^{13}C$ is in contrast reflecting plants activity and their influence on soil making. $\delta^{13}C$ is thus more an indirect variable, being a surrogate of the vegetation type and thus also partly of the interactions with the other species in the local community. Thus soil pH is expected to be more influent on the processes shaping plant distributions. Still, $\delta^{13}C$ showed a greater variable importance than pH when both were added to SDMs. It may be that soil pH interacts with other predictors, resulting in a higher predictive capacity when incorporated in a multi-predictors model.

As expected, stacking individual species predictions (S-SDM) yielded much species richness over-prediction. Adding the two edaphic factors as predictors led to a decrease of the mean species richness over-estimation by 18%. Dubuis et al., (2013) found a similar decrease in over-estimation. We were thus able to confirm their results obtained on 252 plots with our study on 912 plots. Moreover, we only modeled and added two edaphic factors among the 40 edaphic factors measured initially. Some of those factors also showed good results in the exploratory analyses used to select the edaphic factors to be modeled. Future studies should further investigate their potential to predict plant species distributions.

Future perspectives

One additional way to improve $\delta^{13}C$ model fitting could be to include the plant pathway type (C3, C4, CAM) in the predictor set, as done in Suits et al., (2005) study. Furthermore, it is important to notice that our $\delta^{13}C$ model was calibrated using the data from 102 plots only, which may not be enough to capture the whole variation of this factor across our complex study area. Adding more soil samples would likely improve the accuracy of the model. Finally, we used the water-pH method to calculate pH values; another option would have been to calculate pH using the KCl-pH method, as this latter measure usually yields values that are temporally more stable. It represents the potential acidity that a soil can have whereas water pH represents the actual acidity (Gobat et al., 2004).

Conclusion

Plants represent only one component of an ecosystem, so to understand plant distributions one should theoretically also take into account all biotic factors that can drive their distribution. As a result, many recent studies shifted the focus from improving abiotic predictors to adding missing biotic factors in SDMs (Pulliam, 2000; Leathwick & Austin, 2001; Anderson et al., 2002; Lortie et al., 2004; Meier et al., 2010; Wisz et al., 2013). Yet in this study we showed that important abiotic information is still missing in SDMs to capture all abiotic drivers of plant species distributions. Using modeling techniques to generalize edaphic information in space and predict plant species distributions revealed a great potential at the regional scale and complex landscape considered in this study. We therefore suggest that, in future studies, soil information should be more generally considered when building plant SDMs.

Acknowledgment

We thank all the people who helped in the fieldwork, especially V. Buri, M. Scriba, A. Reymond and F. Bouyarden. We are grateful to J.E Spangenberg, T. Adatte, G. Pinard and T. Monnier for their help in the lab and for their advice on soil analysis. Finally, we thank E. Yashiro, L. Desponds, N. Jollien and S. Butta for their support.

References

Aeschimann D. & Burdet H.M. (2008) *Flore de la Suisse et des territoires limitrophes. Le nouveau Binz.* Berne.

Akaike H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Anderson R.P., Peterson A.T., & Gomez-Laverde M. (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.

Andreux F., Cerri C., Vose P.B., Vitorello V.A., F. H.A., Ineson P., & Heal O.W. (1990) Potential of stable isotope, 15N and 13C, methods for determining input and turnover in soils. 259–275.

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

Austin M.P. (1980) Searching for a model for use in vegetation analysis. *Vegetatio*, **42**, 11–21.

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

Baize D., Girard M.-C., & AFES (2009) *Référentiel pédologique 2008.* Editions Quae,

Balesdent J., Mariotti A., Boutton T.W., & Yamasaki S.I. (1996) Measurement of soil organic matter turnover using 13C natural abundance. 83–111.

Battle M., Bender M.L., Tans P.P., White J.W. C., Ellis J.T., Conway T., & Francey R.J. (2000) Global Carbon Sinks and Their Variability Inferred from Atmospheric O2 and 13C. *Science*, **287**, 2467–2470.

Beadle N.. C.. W.. (1953) The Edaphic Factor in Plant Ecology With a Special Note on Soil Phosphates. *Ecology*, **34**, 426–428.

Billings W.D. (1938) The Structure and Development of Old Field Shortleaf Pine Stands and Certain Associated Physical Properties of the Soil. *Ecological Monographs*, **8**, 437–500.

Bishop T.F.A. & McBratney A.B. (2006) A comparison of prediction methods for the creation of field-extent soil property maps. *Geoderma*, **103**, 149–160.

Bolan N.S. & Kandaswamy K. (2005) *Encyclopedia of soils in the environment.* Academic Press, London.

Bouët M. (1972) *Climat et météorologie de la Suisse romande.*

Boutton T.W. & Yamasaki S.I. (1996) Stable carbon isotope ratios of soil organic matter and their use as indicators of vegetation and climate change. 47–82.

Box E.O., Crumpacker D.W., & Hardin E.D. (1993) A Climatic Model for Location of Plant Species in Florida, U.S.A. *Journal of Biogeography*, **20**, 629–644.

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

Cain S.A. (1944) *Foundations of plant geography.* Harper and Brothers, N.Y.,

Castrignanò A., Buttafuoco G., & Comolli R. (2011) Using Digital Elevation Model to Improve Soil pH Prediction in an Alpine Doline. *Pedosphere*, **21**, 259–270.

Chytr M., Tich L., & Role 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. & 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. (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. (2013) Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. *Journal of Vegetation Science*, **24**, 593–606.

Duchaufour P. (1989) Pédologie et groupes écologiques. I: Rôle du type d'humus et du pH. *Bulletin d'écologie*, **20**, 1–6.

Ehleringer J.R., Field C.B., Lin Z., & Kuo C. (1986) Leaf carbon isotope and mineral composition in subtropical plants along an irradiance cline. *Oecologia*, **70**, 520–526.

Ekblad A., Boström B., Holm A., & Comstedt D. (2005) Forest soil respiration rate and delta13C is regulated by recent above ground weather conditions. *Oecologia*, **143**, 136–42.

Engler R., Randin C.F., Vittoz P., Czáka 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.

Evans J., Sharkey T., Berry J., & Farquhar G. (1986) Carbon Isotope Discrimination measured Concurrently with Gas Exchange to Investigate CO 2 Diffusion in Leaves of Higher Plants. *Australian Journal of Plant Physiology*, **13**, 281.

Farquhar G.D., von Caemmerer S., & Berry J.A. (1980) A biochemical model of photosynthetic CO2 assimilation in leaves of C 3 species. *Planta*, **149**, 78–90.

Farquhar G.D., Ehleringer J.R., & Hubick K.T. (1989) Carbon Isotope Discrimination and Photosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology*, **40**, 503–537.

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.

Gillson L., Waldron S., & Willis K.J. (2004) Interpretation of soil δ 13 C as an indicator of vegetation change in African savannas. *Journal of Vegetation Science*, **15**, 339–350.

Gobat J.-M., Aragno M., & Matthey W. (2004) *The Living Soil: Fundamentals of Soil Science and Soil Biology.*

Good R. (1964) *The geography of the flowering plants.* Longmans Green & Co. Ltd., London, Longmans, London.

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

Hastie T.J. & Tibshirani R.J. (1990) *Generalized Additive Models.*

Heegaard E. (2002) A model of alpine species distribution in relation to snowmelt time and altitude. *Journal of Vegetation Science*, **13**, 493–504.

Hengl T., Heuvelink G.B.M., & Stein A. (2004) A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, **120**, 75–93.

Hijmans R.J. & Graham C.H. (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, **12**, 2272–2281.

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

Hossner L. (2008) *Encyclopedia of Soil Science.* Springer, Dordrech.

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

Keeling R.F., Manning A.C., & Dubey M.K. (2011) The atmospheric signature of carbon capture and storage. *Philosophical transactions of the Royal Society A*, **369**, 2113–32.

Körner C. (1999) *Alpine Plant Life.* Springer Berlin Heidelberg, Berlin, Heidelberg.

Körner C., Farquhar G.D., Roksandic Z., & Url S. (1988) A global survey of carbon isotope discrimination in plants from high altitude. *Oecologia*, **74**, 623–632.

Körner C., Farquhar G.D., & Wong S.C. (1991) Carbon isotope discrimination by plants follows latitudinal and altitudinal trends. *Oecologia*, **88**, 30–40.

Lafargue E., Espitalié J., Marquis F., & Pillot D. (1998) Rock-Eval 6 Applications in Hydrocarbon Exploration, Production, and Soil Contamination Studies. *Revue de l'Institut Français du pétrole*, **53**, 421–437.

Leathwick J.R. & Austin M.P. (2001) COMPETITIVE INTERACTIONS BETWEEN TREE SPECIES IN NEW ZEALAND'S OLD-GROWTH INDIGENOUS FORESTS. *Ecology*, **82**, 2560–2573.

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.

Lozet J. & Mathieu C. (1997) *Dictionnaire de Science du Sol.*

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**, 1933–9.

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

McCullagh P. & Nelder J.A. (1989) *Generalized Linear Models, Second Edition.*

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.

Nagelkerke N.J.D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, **78**, 691–692.

Ninyerola M., Pons X., & Roure J.M. (2000) A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *International Journal of Climatology*, **20**, 1823–1841.

O'Leary M.H. & Osmond C.B. (1980) Diffusional Contribution to Carbon Isotope Fractionation during Dark CO2 Fixation in CAM Plants. *PLANT PHYSIOLOGY*, **66**, 931–934.

Page A.L. (1982) *Methods of soil analysis. Part 2. Chemical and microbiological properties.* American Society of Agronomy, Soil Science Society of America,

Park S.J. & Vlek P.L.. (2002) Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques. *Geoderma*, **109**, 117–140.

Pärtel M. (2002) Local plant diversity patterns and evolutionary history at the regional scale. *Ecology*, **83**, 2361–2366.

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**, 1805–1813.

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

Pradervand J.-N., Dubuis A., Pellissier L., Guisan A., & Randin C. (2013) Very high resolution environmental predictors in species distribution models: Moving beyond topography? *Progress in Physical Geography*, .

Pulliam H.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. (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. (2009) Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology*, **15**, 1557–1569.

Roem W.J. & Berendse F. (2000) Soil acidity and nutrient supply ratio as possible factors determining changes in plant species diversity in grassland and heathland communities. *Biological Conservation*, **92**, 151–161.

Shao G. & Halpin P.N. (1995) Climatic Controls of Eastern North American Coastal Tree and Shrub Distributions. *Journal of Biogeography*, **22**, 1083–1089.

Smith B.N. & Epstein S. (1971) Two Categories of 13C/12C Ratios for Higher Plants. *PLANT PHYSIOLOGY*, **47**, 380–384.

Soil Survey Staff (2009) *Soil Survey Field and Laboratory Methods Manual. Soil Survey Investigations Report No. 51, Version 1.0.*

Spangenberg J.E., Jacomet S., & Schibler J. (2006) Chemical analyses of organic residues in archaeological pottery from Arbon Bleiche 3, Switzerland – evidence for dairying in the late Neolithic. *Journal of Archaeological Science*, **33**, 1–13.

Stewart G., Turnbull M., Schmidt S., & Erskine P. (1995) 13 C Natural Abundance in Plant Communities Along a Rainfall Gradient: a Biological Integrator of Water Availability. *Australian Journal of Plant Physiology*, **22**, 51.

Suits N.S., Denning a. S., Berry J. a., Still C.J., Kaduk J., Miller J.B., & Baker I.T. (2005) Simulation of carbon isotope discrimination of the terrestrial biosphere. *Global Biogeochemical Cycles*, **19**, .

Tamburini F., Adatte T., Föllmi K., Bernasconi S.M., & Steinmann P. (2003) Investigating the history of East Asian monsoon and climate during the last glacial–interglacial period (0–140 000 years): mineralogy and geochemistry of ODP Sites 1143 and 1144, South China Sea. *Marine Geology*, **201**, 147–168.

Thuiller W., Georges D., & Engler R. (2013) biomod2: Ensemble platform for species distribution modeling. R package version 3.1-18. http://CRAN.R-project.org/package=biomod2. .

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.

Vogel J.C. (1980) *Fractionation of the Carbon Isotopes During Photosynthesis.* Springer Berlin Heidelberg, Berlin, Heidelberg.

Vogel J.C. (1993) Variability of Carbon Isotope Fractionation during. *Stable isotopes and plant carbon-water relations*, 29.

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., Öckinger 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.

Woodward F.I. (1987) *Climate and Plant Distribution.*

Figures

Figure 1

Map of the study area located on the western Swiss Alps. The 242 plots for which we have edaphic information are symbolized; squares represent the 102 plots sampled in 2012 and circles the 140 plots sampled in 2013.

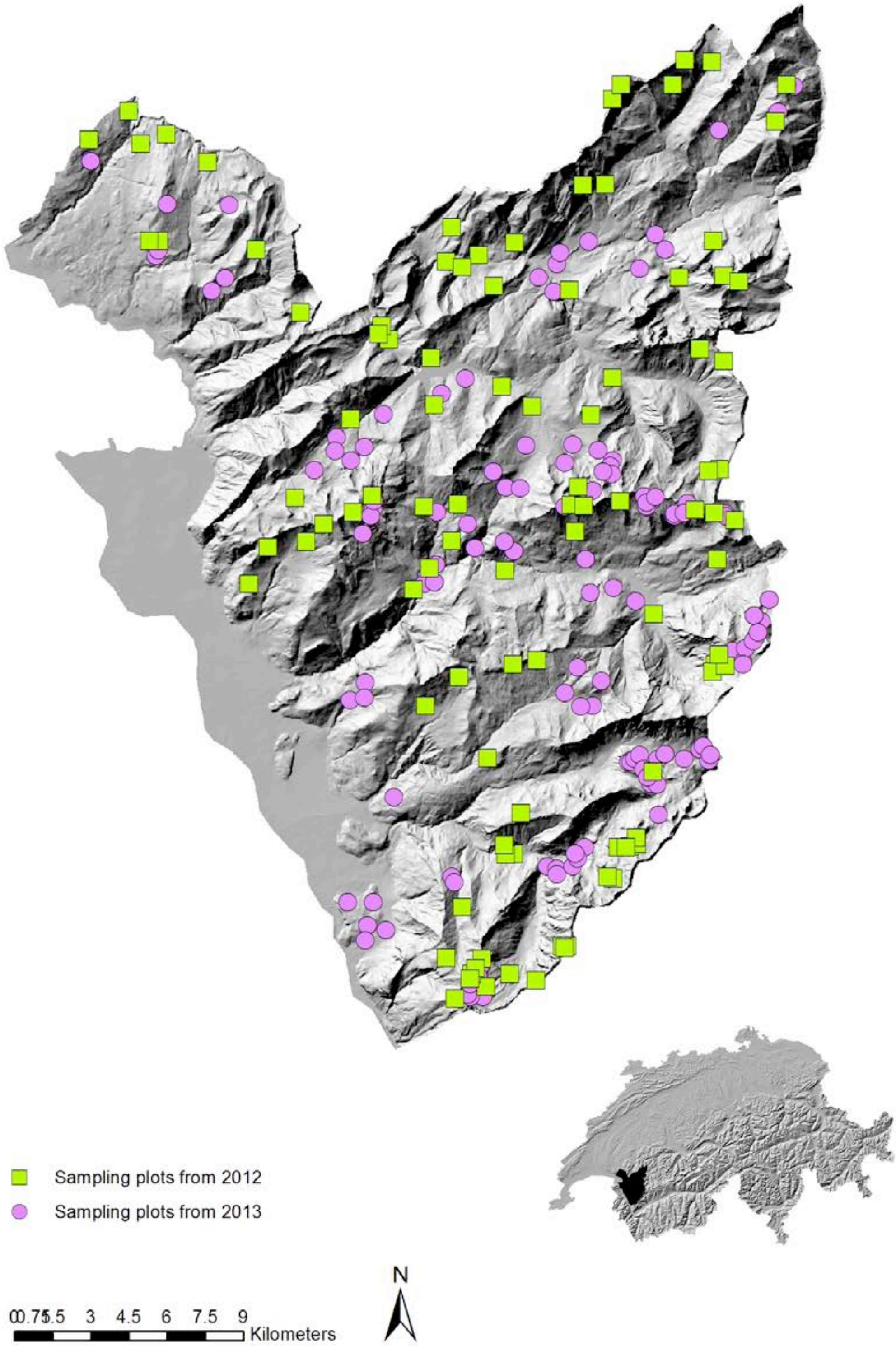


Figure 2

Importance of the edaphic variables. The importance calculated by Biomod2 (mean for each modeling techniques) of the 40 edaphic variables used as predictors in the model for 42 plant species is represented ranked by increasing order. The boxplot limits are the 1st and the 4th quartiles and the inner line represents the median.

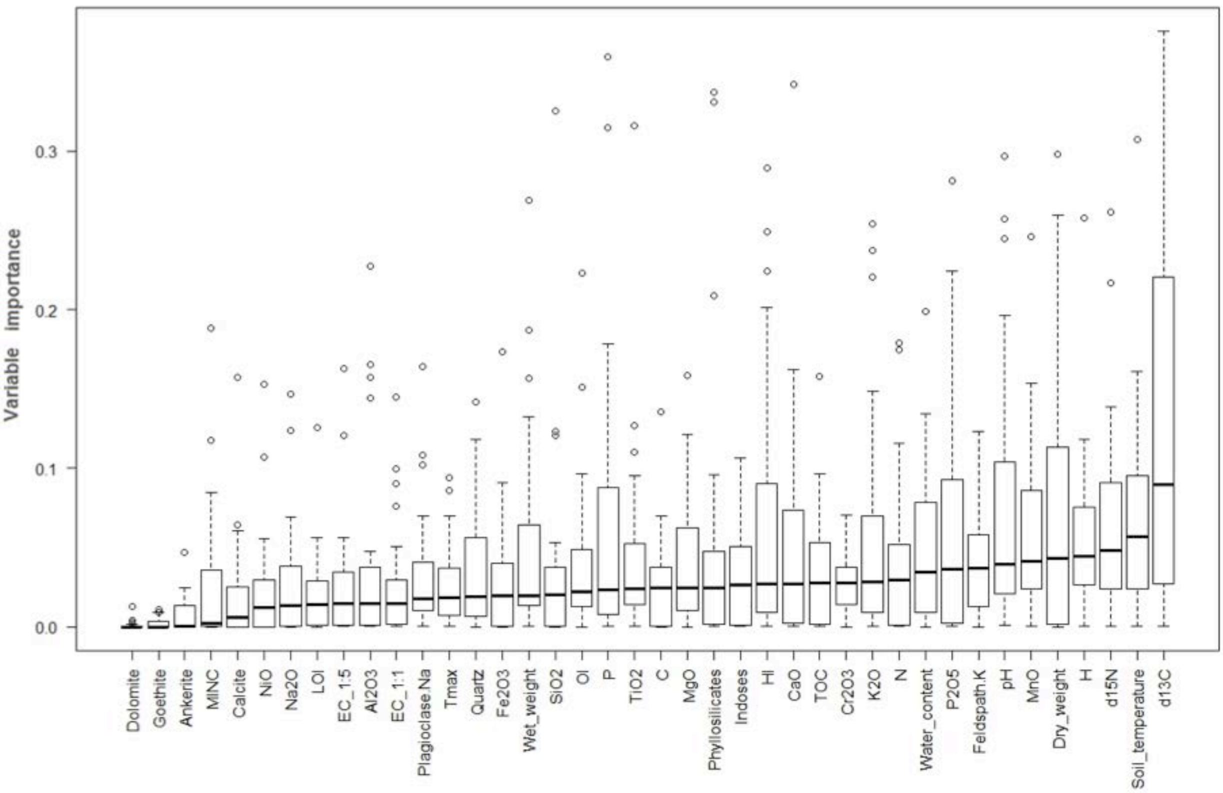


Figure 3

Map of the predicted $\delta^{13}C$ for the western Swiss Alps. A the bottom right is a histogram of the distribution of the predicted $\delta^{13}C$.

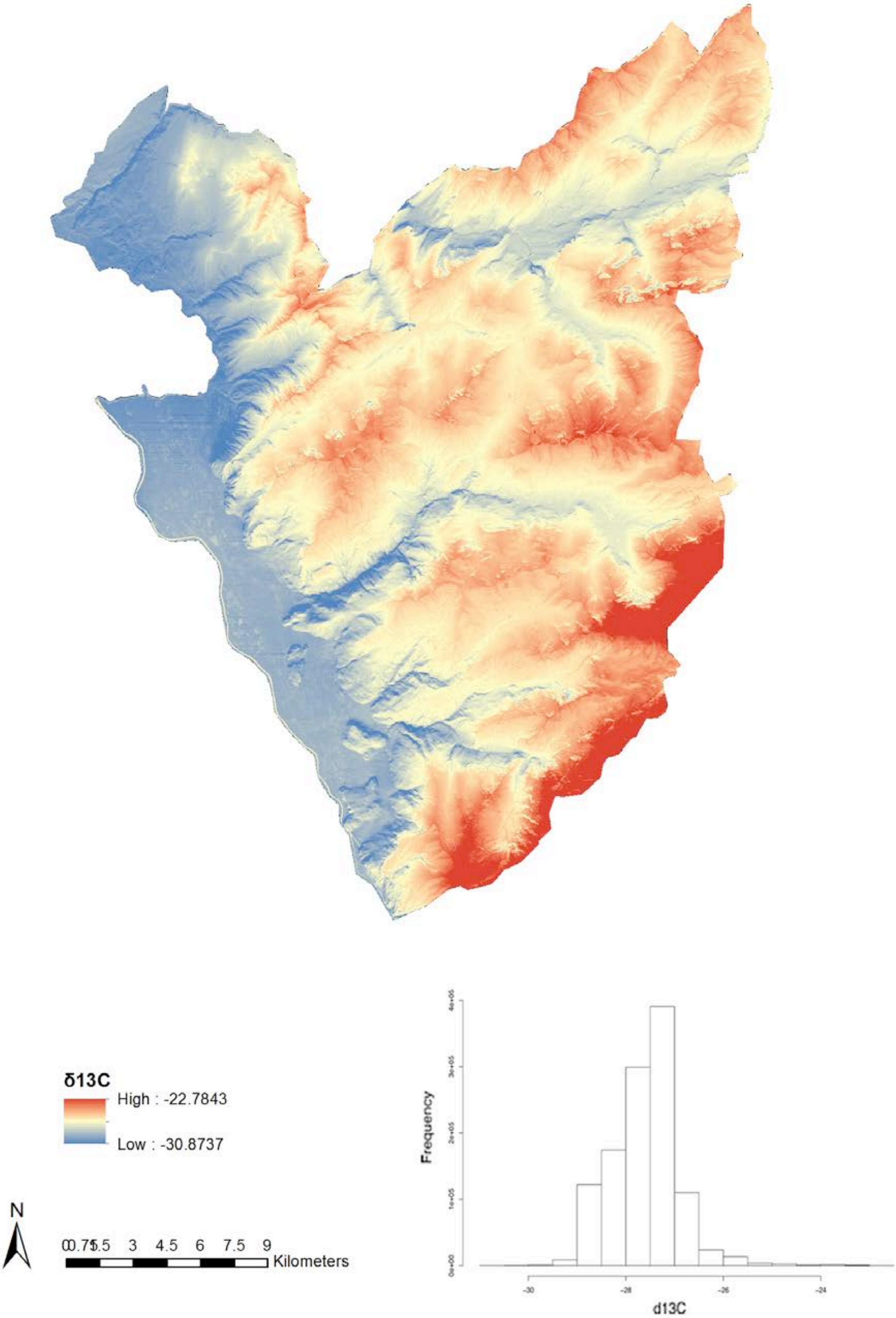


Figure 4

Map of the predicted pH for the western Swiss Alps. A the bottom right is a histogram of the distribution of the predicted pH .

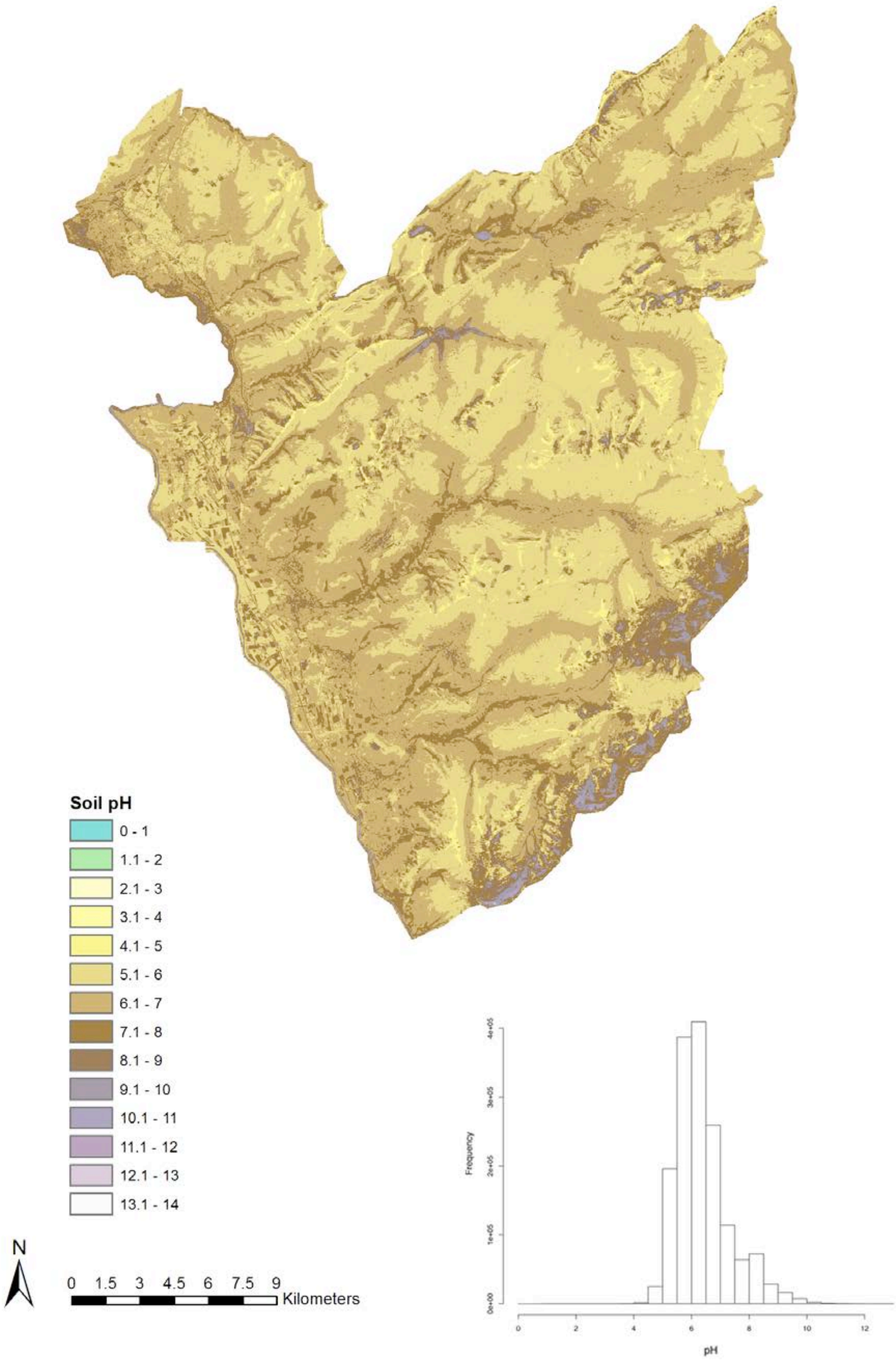


Figure 5

Importance of the variables used as predictors in the four models types. The importance calculated by biomod2 (mean for each modeling techniques) in the model for 260 plant species with the use of (a) the 4 topo-climatic variables (4TC) as predictors, (b) the 4TC and the $\delta^{13}C$ as predictors, (c) the 4TC and the pH as predictors, (d) the 4TC, the $\delta^{13}C$ and the pH as predictors. Blue color represents the $\delta^{13}C$ importance and red color represents the soil pH importance. The boxplot limits are the 1st and the 4th quartiles and the inner line represents the median.

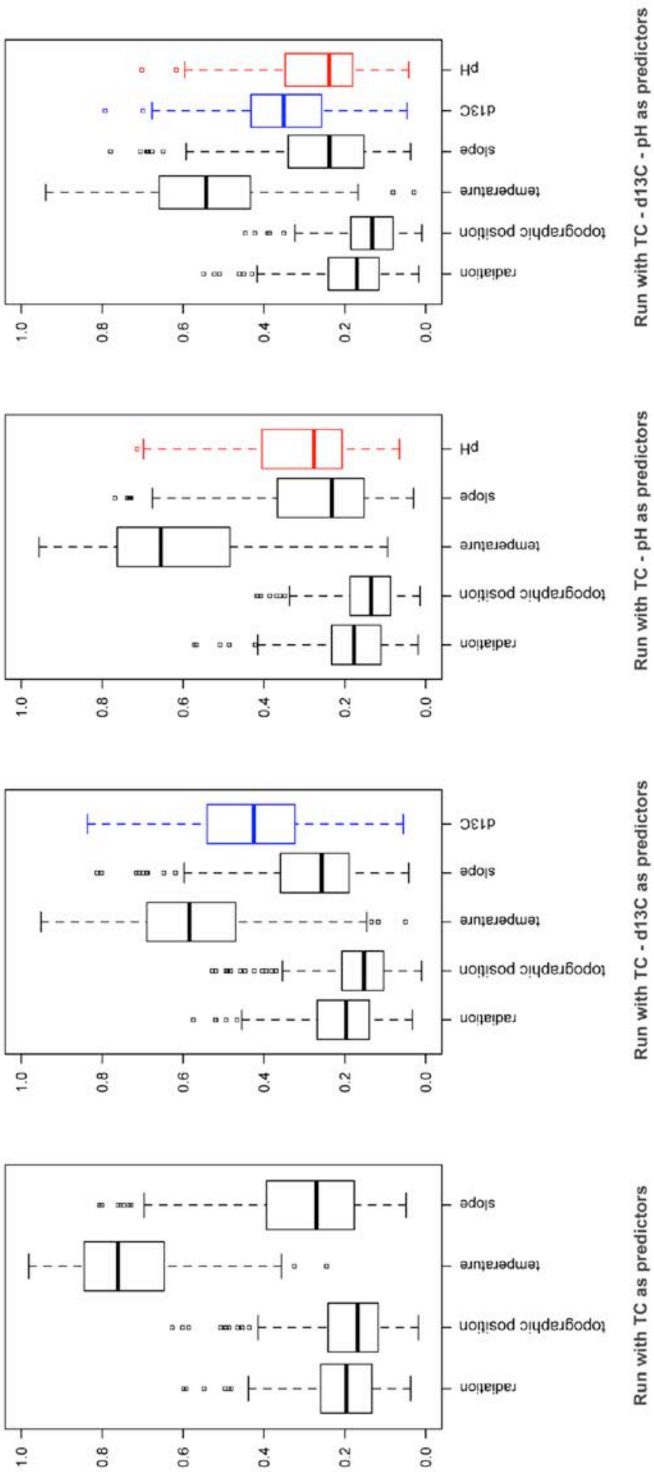


Figure 6

Boxplots of the area under the curve of a receiver-operating characteristic plot (AUC) for the evaluations of the four models types (using as predictor either four topo-climatic variables, four topo-climatic variables + $\delta^{13}C$, four topo-climatic variables + soil pH or four topo-climatic variables + soil pH + $\delta^{13}C$) for each species.

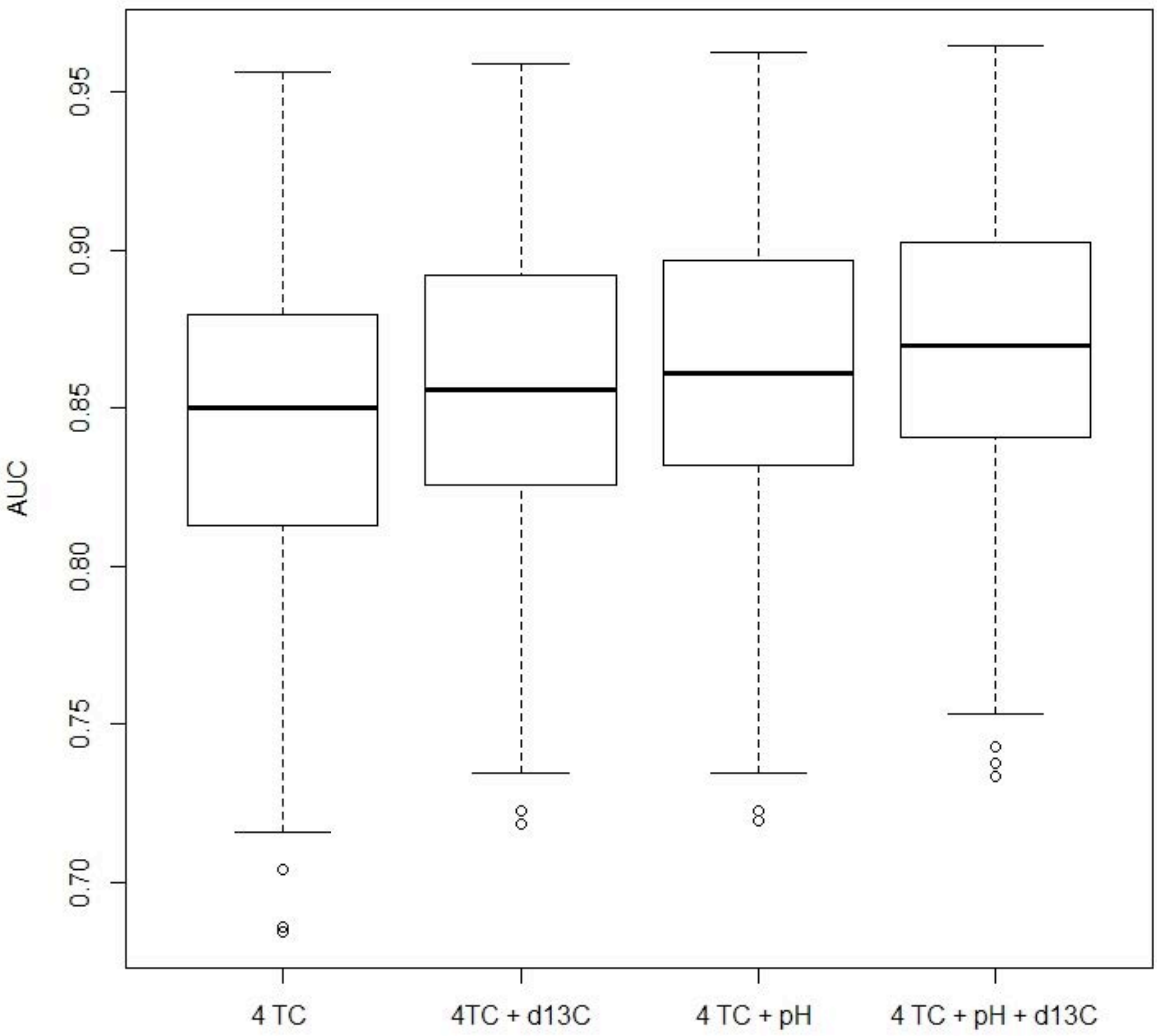
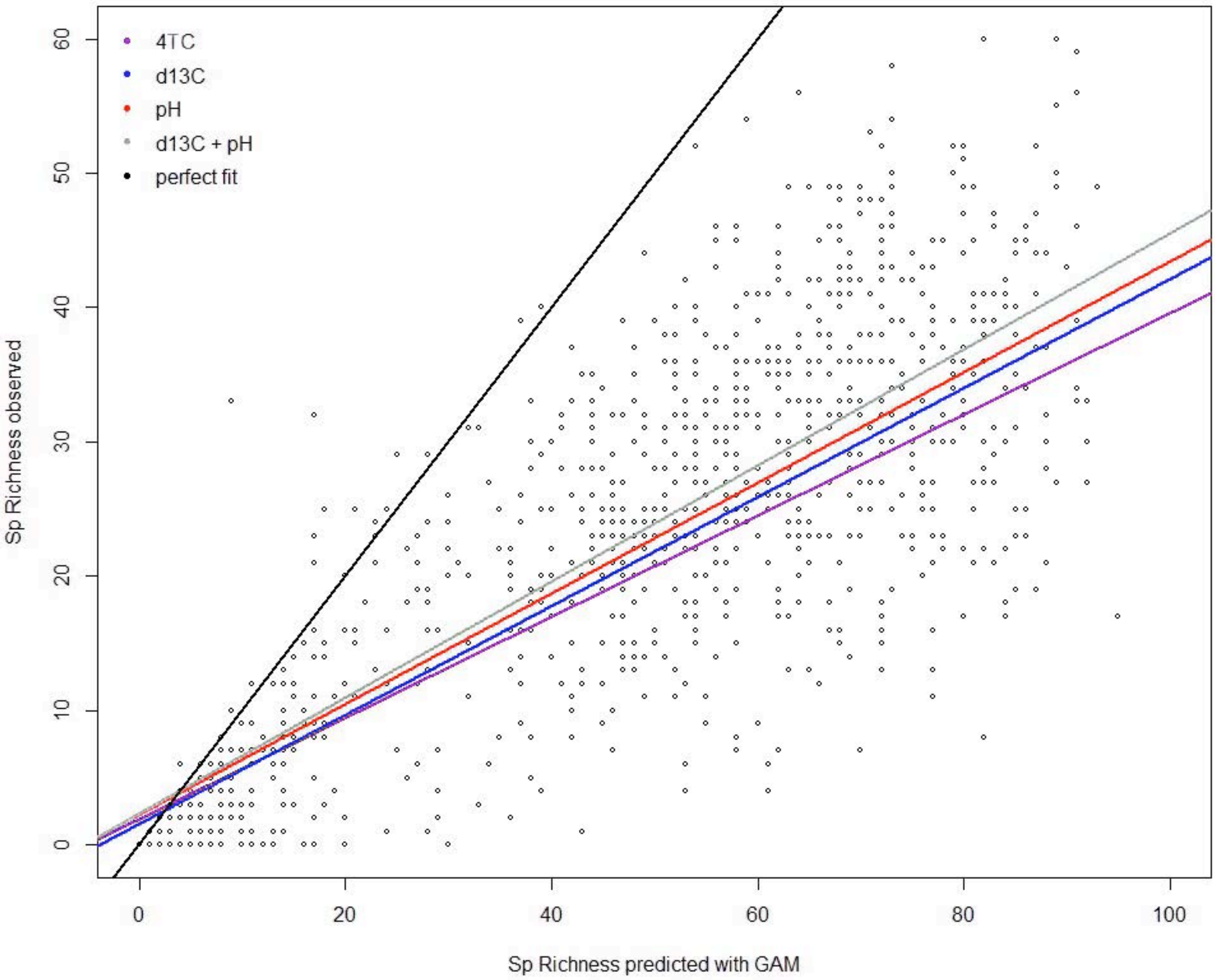


Figure 7

Predicted species richness plotted against the observed data. The violet line corresponds to the linear regression for the S-SDM using the 4TC as predictors. The blue line is related to the one using 4TC and $\delta^{13}C$. The red line corresponds to the one using ATC and soil pH. The grey line is linked to the one using 4TC, $\delta^{13}C$ and soil pH. The black line represents a perfect correspondence between predicted and observed species richness.



Tables

Table 1

Potential predictors used in the stepwise selection to model edaphic factors

Variables	Units	Details	Source
Precipitations (average of monthly values for each season)	mm*month-1	Seasonal average of precipitations	MeteoSuisse/WSL
Temperature (average of monthly values for each season)	celcius degrees	Seasonal average of temperatures	MeteoSuisse/WSL
Moisture Index (average of monthly values for each season)	mm day -1	Atmosheric humidity balance over a landscape	WSL
Solar radiation	KJ/day	Daily average of global potential shortwave radiation per month	WSL
Normalized Difference Vegetation Index (NDVI)	unitless	Combination of wavelengths of visible and near-infrared	Swiss topo / NPOC
Normalized Difference Snow Index (NDSI)	unitless	Combination of wavelengths of visible and short wave-infrared	Swiss topo / NPOC
Topographic wetness index (TWI)	unitless	Wetness according to slope and drained area	WSL
Site water balance (SWB)	mm*year-1	Annual average site water balance accounting for soil proprieties	WSL
Geotechnical map	rock, soil categories	Soil suitability map	GEOSTAT
Slope	degrees	Slope inclination	OFT - MNT25
Elevation digital model	meters	Elevation	OFT - MNT26
Topographic position	unitless	Relative topographic exposure	WSL

Table 2

For each edaphic factor, this table shows the units, the minimum and the maximum values as well as the average value obtained after chemical analyses.

Edaphic factor	Units	Minimum values	Maximum values	Average
Soil temperature	degree celcius	5.60	24.00	15.57
Soil wet weight	g	189.42	488.99	232.62
Soil dry weight	g	27.70	441.77	159.81
Soil water content	%	4.45	656.10	61.79
pH	unitless	4.37	9.04	6.44
EC_1:5	$\mu S*cm^{-1}$	44.00	638.00	126.34
EC_1:1	$\mu S*cm^{-1}$	220.00	3190.00	631.70
P	mg/g	0.20	2.77	0.80
TOC	%	0.26	40.23	6.43
MINC	%	0.24	10.53	1.59
HI	mg HC* g TOC^{-1}	55.68	461.05	221.38
OI	mg CO2*g TOC^{-1}	153.91	485.72	237.74
Tmax	degree celcius	260.29	427.00	382.94
N	%	0.03	2.67	0.53
C	%	1.29	41.63	7.98
H	%	0.00	4.43	1.26
Phyllosilicates	%	0.00	61.08	35.40
Quartz	%	0.00	63.89	35.13
Feldspath-K	%	0.00	6.46	0.74
Plagioclase-Na	%	0.00	13.86	2.82
Calcite	%	0.00	75.13	7.30
Dolomite	%	0.00	9.65	0.12
Goethite	%	0.00	4.11	0.23
Ankerite	%	0.00	5.05	0.21
Indoses	%	2.25	100.00	18.04
SiO2	wt-%	3.00	79.83	54.62
TiO2	wt-%	0.05	0.86	0.50
Al2O3	wt-%	0.85	17.22	9.45
Fe2O3	wt-%	0.47	7.15	3.84
MnO	wt-%	0.02	0.34	0.09
MgO	wt-%	0.42	3.58	1.30
CaO	wt-%	0.09	51.06	6.16
Na2O	wt-%	0.05	1.30	0.56
K2O	wt-%	0.09	3.52	1.57
P2O5	wt-%	0.05	0.64	0.20
OM	wt-%	6.88	63.69	20.96
Cr2O3	wt-%	0.00	0.04	0.02
NiO	wt-%	0.00	0.01	0.01
$\delta^{15}N$	‰ AIR	-3.10	5.30	1.22
$\delta^{13}C$	‰V-PDB	-29.50	-24.37	-27.10

Table 3

Wilcoxon signed rank test between the AUC of the four models types (using as predictor either four topo-climatic variables (4TC), 4TC + $\delta^{13}C$, 4TC + soil pH or 4TC + soil pH + $\delta^{13}C$).

Tests with significant p-values ($P < 0.05$) are indicated in bold marked with a star.

	Wilcoxon test	W	p-value
GLM	4TC+pH vs 4TC	31162	0.12
	4TC+pH vs 4TC+d13C	35238	0.40
	4TC+pH vs 4TC+pH+d13C	33166.5	0.71
	4TC+d13C vs 4TC	32718.5	0.53
	4TC+d13C vs 4TC+pH+d13C	36035.5	0.19
	4TC+pH+d13C vs 4 TC	30372	**0.05 ***
RF	**4TC+pH vs 4TC**	25762.5	**2.72E-06 ***
	4TC+pH vs 4TC+d13C	36039	0.191
	4TC+pH vs 4TC+pH+d13C	29464	**0.011 ***
	4TC+d13C vs 4TC	28012.5	**7.30E-04 ***
	4TC+d13C vs 4TC+pH+d13C	40170	**2.01E-04 ***
	4TC+pH+d13C vs 4 TC	22047.5	**6.90E-12 ***
GAM	**4TC+pH vs 4TC**	28013.5	**7.32E-04 ***
	4TC+pH vs 4TC+d13C	35377	0.36
	4TC+pH vs 4TC+pH+d13C	30121	**0.03 ***
	4TC+d13C vs 4TC	29910	**0.02 ***
	4TC+d13C vs pH+4TC+d13C	39080	**2.06E-03 ***
	4TC+pH+d13C vs 4 TC	24850	**0.0000001752 ***

Table 4

Percentage of the mean species richness over-estimation of the four models type (using as predictor either four topo-climatic variables (4TC), 4TC + $\delta^{13}C$, 4TC + soil pH or 4TC + soil pH + $\delta^{13}C$) compared to observed species richness. Over-estimation is calculated for GLM, GAM and RF modeling techniques as well as for the mean of the 3 techniques.

	Percentage of over-estimation			
	GLM	**GAM**	**RF**	**ALL**
4 TC	204.6	144.4	101.8	150.0
4 TC + d13C	195.1	130.2	105.8	143.7
4 TC + pH	190.0	119.9	98.2	136.1
4 TC + d13C + pH	184.4	109.2	101.7	131.8

Supporting information

Appendix S1

Table presenting a short description of every analysis used as well as a information about edaphic factors obtained.

Analysis	Description	Information obtained and biological meaning
C and N stable isotopes	Determination of the C and N isotopic ratios ($^{13}C/^{12}C$ and $^{15}N/^{15}N$)	Insight into the main organic contribution (plant debris, microbial biomass have different $\delta^{13}C$, $\delta^{15}N$ values and C/N ratios) and trophic level.
CHN	Determination of C, H and N content; C/N and C/H ratio	Lipids are rich in H compared to other organic compounds. Soil lipids include carboxylic acids (fatty acids) and wax alcohols, n-alkanes, and sterols (mainly from epicuticular waxes). Organic N occur mainly in amino acids (proteins) and are part of main nutrients.
P	Determination of total P	P is one of the main nutrients
Rock-Eval	Determination of parameters during the pyrolysis of soil ornaic matter: hydrogen index (HI), oxygen index (OI), temperature of maximal pyrolytic yield (T-max). Organic and mineral carbon (MINC and TOC)	HI and OI give some information of the SOM (soil organic matter). T-max may give some information on the degree of recalcitrance/polymerization of the SOM
Water pH	Determination of soil pH	Soil pH is important for plant growth and control nutrients supply
Electro Conductivity analysis	Measure the quantity of electrical current a soil can conduct	Inform indirectly about the soil texture
Soil temperature	Measure of the soil temperature at 5 cm deep	Soil temperature influenced soil formation process and life inside the soil as well as on its surface.
Soil gravimetric water content	Ratio from water mass and bulk mass. Wet and dry weight are calculated in this purpose	Inform about the water regime of the soil
X ray fluorescence	Measurement of the soil major elements (SiO_2, TiO_2, Al_2O, Fe_2O, MnO, MgO, CaO, Na_2, K_2O, P_2O, Cr_2O_3, NiO)	Inform about the presence of macro element necessary for plant growth and development
X ray diffraction	Identification of minerals as Phyllosilicate, Quartz, Feldspath-K, Plagioclase-Na, Calcite, Dolomite, Goethite, Ankerite, Indoses in soil.	Inform about soil mineralogy which influence the physical and chemical properties of soils and can be uses in soil classification

Appendix S2

Variables used as a basic set of topo-climatic predictors (4TC) in plant species distribution modeling.

Variables	Units	Details	Source
Summer air temperature	celcius degrees	Average of temperatures for June to August	MeteoSuisse/WSL
Summer radiation	KJ/day	Daily average of global potential shortwave radiation for June to August	WSL
Slope	degrees	Slope inclination	OFT - MNT25
Topographic position	unitless	Relative topographic exposure	WSL

Appendix S3

ΔR^2 of the different factors used to model soil pH and $\delta^{13}C$.

	Predictor	ΔR^2
Soil pH	winter moisture index	0.0364
	spring radiation	0.0152
	normalized difference vegetation index	0.0162
	normalized difference snow index	0.0178
	site water balance	0.0143
	topographic position	0.0397
$\delta^{13}C$	winter temperature	0.0006
	autumn moisture index	0.0149
	spring radiation	0.0122
	normalized difference snow index	0.0076
	elevation	0.0160