Spatial modelling of soil water holding capacity improves models of plant distributions in mountain landscapes

Cianfrani C.[1,*,#], Buri A.[1,*], Vittoz P.[1], Grand S.[1], Zingg B.[1], Verrecchia E.[1], Guisan A.[1,2]

[1]Institute of Earth Surface Dynamics (IDYST), University of Lausanne (UNIL), 1015 Lausanne, Switzerland
[2]Department of Ecology and Evolution (DEE), University of Lausanne (UNIL), 1015 Lausanne, Switzerland
[#]Corresponding author: carmen.cianfrani@unil.ch
[*]Equally contributed

Abstract

Aims

The aims of this study were: 1) to test a new methodology to overcome the issue of the predictive capacity of soil water availability in geographic space due to measurement scarcity, 2) to model and generalize soil water availability spatially to a whole region, and 3) to test its predictive capacity in plant SDMs.

Methods

First, we modelled the measured Soil Water Holding Capacity (SWHC at different pFs) of 24 soils in a focal research area, using a weighted ensemble of small bivariate models (ESM). We then used these models to predict 256 locations of a larger region and used the differences in these pF predictions to calculate three different indices of soil water availability for plants (SWAP. These SWAP variables were added one by one to a set of conventional topo-climatic predictors to model 104 plant species distributions.

Results

We showed that adding SWAP to the SDMs could improve our ability to predict plant species distributions, and more specifically, pF1.8–pF4.2 became the third most important predictor across all plant models.

Conclusions

Soil water availability can contribute a significant increase in the predictive power of plant distribution models, by identifying important additional abiotic information to describe plant ecological niches.

Key words: ensemble of small weighted bivariate models, soil water holding capacity, habitat suitability, predictions, environmental niche, topo-climate, Swiss Alps

Introduction

The capacity of soils to retain water is known to be an important factor that influences plant distributions, with effects on carbon allocation, microbial activity, nutrient cycling, and the photosynthetic rate (Adhikari & Hartemink 2016). Under climate change, soil water content is expected to suffer from the decreasing summer precipitation and air humidity and the increasing solar radiation and evapotranspiration (Jasper, Calanca & Fuhrer 2006). This decrease in soil water content added to the predicted increase in the severity and frequency of drought events will result in more stressful water conditions for plants (Calanca 2007). Thus, the capacity of the soils to retain water might become critical for supplying water to mountain plants (Jasper, Calanca & Fuhrer 2006), which was already shown to affect plant species diversity (Kammer *et al.* 2013).

Despite its importance, the soil capacity to retain water was only used as an explanatory variable in a very few interdisciplinary studies, such as vegetation modelling (e.g., Mod *et al.* (2016), likely due to the paucity of available data caused by the time-consuming and expensive laboratory analyses required to obtain them. This paucity also prevented its spatial generalization to large areas or entire regions, and consequently, its inclusion in field-based spatially explicit modelling studies. The potential benefit of including soil water content in plant species distribution models (SDMs) has already been argued (Guisan & Zimmermann 2000; Austin & Van Niel 2011; Adhikari & Hartemink 2016; Mod *et al.* 2016) because it provides a way to determine the soil moisture content necessary for the growth of a given species (Viji & Rajesh 2012). Mod *et al.* (2016) pointed out that the variables that are generally used in SDMs to represent water availability for plants are derived from precipitation, but the latter does not measure the amount of water actually available for plants. Consequently, the real measurement of the soil water available for plants is rarely considered to be an explanatory variable in plant distribution modelling.

Soil water availability may be estimated using the soil water retention curve, which relates the soil water content to applied suction or tension (Buckman & Brady 1922; Richards & Weaver 1944). Tension is represented by the matric potential (F), which corresponds to the force necessary to extract water from soils, thereby overcoming capillary retention. The logarithm of the absolute value of matric potential is commonly used as a measure of applied tension (Gobat, Aragno & Matthey 2004):

$$pF = \log_{10} |F| \quad (1)$$

with F expressed as the height of the water column (cm).

The soil water content at different values of pF characterizes the soil water holding capacity (SWHC). The following points may be of interest on the water retention curve. For instance, water content at pF=4.2, which is named the permanent wilting point, corresponds to the point where most plants will wilt beyond recovery. Water content at pF=3.7, which is sometimes named the temporary wilting point, corresponds to the point where most plants will begin to experience severe water stress. Water content at pF=2.7, which is named the field capacity, is defined as the amount of water remaining in the soil after downward gravitational drainage has stopped, i.e., the water retained naturally in the soil (Fig. 1).

The change in water content between different pFs is used to estimate the quantity of water that may be available for plant uptake in soil. Plant available water is conventionally defined as the difference between the field capacity and the permanent wilting point. It is also possible to define other indices of water availability. For instance, the difference between the field capacity and the temporary wilting point represents the range of optimal hydric status for most plants, while the difference between the field capacity for sands and the permanent wilting point could be an upper-bound indicator of accessible water in soils, including coarse-textured soils, which have often lost most of their water by pF=2.7 (Veihmeyer & Hendrickson 1928) (Fig. 1).
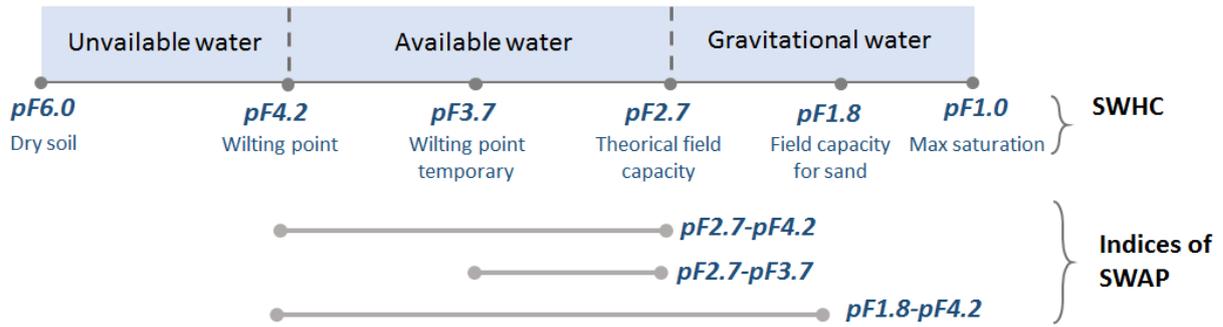
Fig. 1 – Types of soil water availability in relationship to its corresponding points on the water retention curve. The soil water content at different values of pF characterizes the soil water holding capacity (SWHC). The change in water content between different pFs is used to estimate the quantity of water that may be available for plant uptake in soils (indices of SWAP).

The field capacity and wilting point values can change depending on the environment and the plants species, making different soils and habitats react differently. However, we still have very limited knowledge on the spatial variability of these properties in natural ecosystems, especially in mountain areas. Numerous studies have focused on the development of pedotransfer functions, which aim to predict the SWHC using basic soil properties. In particular, during the last two decades, continuous pedotransfer functions have been developed (Wosten, Pachepsky & Rawls 2001; Patil & Singh 2016). However, their prediction accuracy depends on the number of soil samples used for their calibration, although these are often insufficient for developing a good model, especially in a heterogeneous non-agricultural landscape (De Vos *et al.* 2005).

In this study, we aimed to solve the issue of data scarcity of SWHC measurements in hampering prediction accuracy, by investigating the effect of soil water holding capacity on plant species distributions across a whole temperate mountain region. For this purpose, we: i) used a new modelling method to spatially predict local SWHC measurements over an entire region, ii) used the predicted SWHC to calculate three indices of soil water available for plants (SWAP) obtained by the differences among the SWHC values, and iii) included the SWAP indices as

predictors in plants species distribution models (plant SDMs) in order to assess if their inclusion resulted in improvements of the plant SDMs.

Materials and methods

General study framework

This study was based on an intensive sampling of a sub-region (NAA sector) of our larger study area in order to test the capacity to predict the parameter values for the whole region (VA area). The study was carried out following four different steps (Fig. 2). The first step consisted of using a new method for low sample sizes to spatially predict local SWHC measurements (24 measurements in the NAA sector at pF=1.8, pF=2.7, pF=3.7, pF=4.2; see Fig. 2) over larger areas (256 locations in the whole VA area; see Fig. 2). The second step consisted of using the predicted SWHC to calculate three values of SWAP obtained by the differences among the SWHCs *([pF2.7-pF4.2]*, *[pF1.8-pF4.2]* and *[pF2.7-pF3.7]*). The third step consisted of the inclusion of three indices of SWAP as predictors in plants species distribution models (plant SDMs) over the larger region, for which large species observation data sets were already available. Finally, the last steps aimed to assess whether the inclusion of SWAP improved the plant SDMs (Fig. 3).

Study area

The study was conducted in the Swiss western Alps covering the Vaud Canton (Vaud Alps, VA in Fig. 2). The study area covered approximately 700 km^2 and encompassed whole massifs, thus spanning a large elevation gradient, from 375 to 3210 m a.s.l. Natural vegetation changes along the elevation gradient, starting from scarce nival vegetation patches at high elevations, to alpine grasslands, to subalpine heaths and coniferous forests, to mixed coniferous-broadleaf forests at mid elevation (montane), to broadleaf forests at low elevation. The mountainous areas

(montane to subalpine) are under moderate human influence and are mainly used for mowing and grazing by sheep and cattle that maintain many open habitats; however, human influence is quite high in the lowlands, where more intense use of the land and urbanization are observed (see http://rechalp.unil.ch). SWHC measurements were obtained from the soil samples collected in the Vallon de Nant-Argentine-Anzeindaz sector (NAA sector in Fig. 2), which is located in the southwestern part of the larger VA area and ranges from mid- to high elevations.
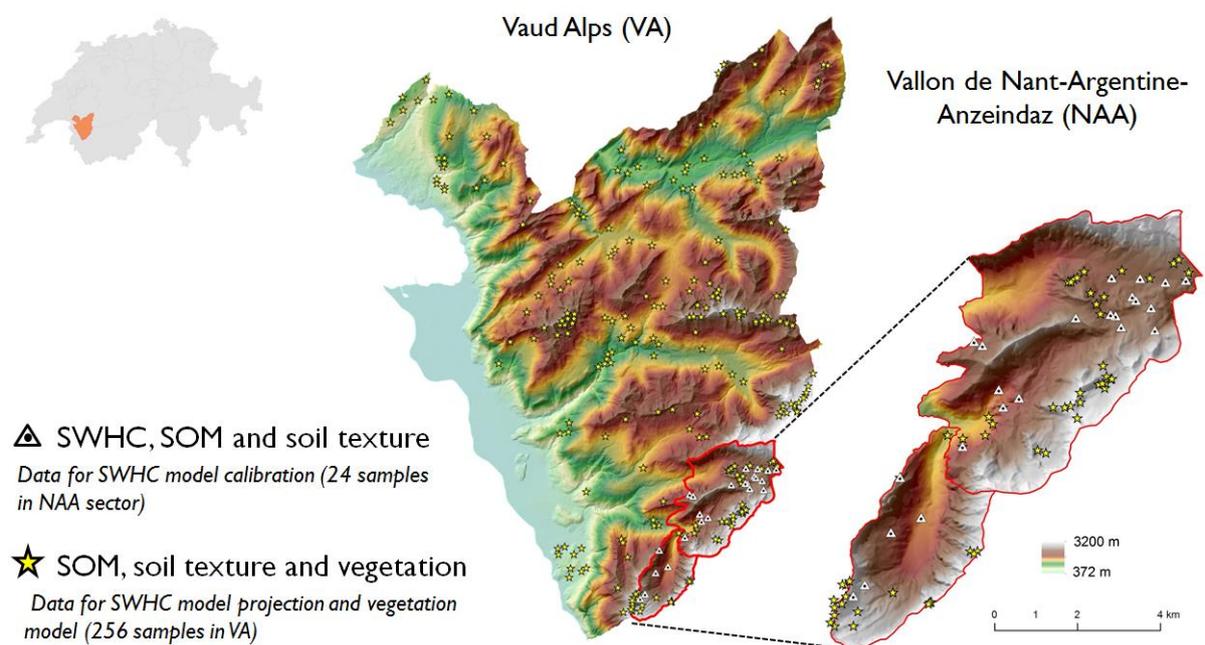

Fig. 2 – Study area and sampling strategy. SWHC: soil water holding capacity, SOM: soil organic matter.

Step 1 - Soil water holding capacity (SWHC) modelling through weighted Ensemble of Small Models (ESM)

SWHC measurements

Twenty-four soil samples were collected in the NAA sector during the summer 2014 by following an equal random-stratified design (Hirzel & Guisan 2002), with the elevation, slope

and vegetation type (Randin *et al.* 2009b; Dubuis *et al.* 2013; Buri *et al.* 2017) as stratifying factors to take into account the variability of the study area. Soil samples were extracted from the organo-mineral topsoil (5-10 cm deep) by taking off a core of circa 4 x15 cm and 350 g of soil for analysis after drying at 105°C. The SWHC was measured in the laboratory at four different matrix potentials or pressures that correspond to different water availabilities for plants: pF=1.8, pF=2.7 (field capacity), pF=3.7 and pF=4.2 (wilting point; see Fig. SI1, Fig. SI2 and Tab. SI1 in Supplementary Information for measured values). The soil volumetric water content (cm^3/cm^3) was determined using a sandbox (Eijkelkamp, Sandbox 08.01) for the lower pF values (i.e. < pF 2.5) and a ceramic plate extractor for the higher pF values (Soil Moisture Equipment Corp, ceramic plate extractor 1500). Soil samples were first saturated with water, then different pressures were applied (through suction or constant pressure) and measures were taken once the equilibrium with the applied pressure was reach. Finally, samples were oven dried at 105°C to find their dry weight which were used to calculate the final volumetric water content at each pF.

Soil texture and organic matter measurements

Soil texture and organic matter content were measured for 24 samples of the NAA sector during summer 2014 and at 256 locations in the VA area during the summers of 2012 and 2013, with the following sampling method: a 4 m^2 square was delimited for each plot, and the soil cores were extracted from the topsoil (5-10 cm deep) at the four corners and in the centre of the square, and the five samples were mixed and homogenized. We selected samples by following an equal random-stratified design that is particularly suited for fitting species distribution models (Hirzel & Guisan 2002), with the elevation, slope and aspect as stratifying factors. Soil samples were first dried and sieved at 2 mm (for fine earth). Then, organic matter (OM) was removed from a subsample of the soil by adding 10-35% hydrogen peroxide (H_2O_2). This

fraction of the soil was then analysed by laser diffraction (Malvern™ Mastersizer 2000) to measure the particle size distribution. Textural classes were attributed according to the USDA texture classification (Shirazi & Boersma 1984).

For the NAA sector, a loss of ignition (LOI) at 450°C was performed to determine the OM content. For the VA sector, we characterized the total organic carbon (TOC) by thermal analysis on crushed samples using a Rock-Eval 6 Pyrolyser (Vinci Technologies) with a standard whole-rock pyrolysis method (Lafargue, Marquis & Pillot 1998). Finally, we used the formula TOC=SOM / 1.724 to convert the result of LOI to TOC (Pribyl 2010).

Ensemble of small models of SWHC

We fitted models for each of the 4 SWHC pF measured in the 24 plots using the R-3.5.0 software. The soil organic matter (TOC), soil texture (proportion of sand, silt, clay), and topographic variables (elevation, aspect and convexity) were used as predictors (Fig. 2; Tab. SI2 in Supplementary Information). The TOC, sand, silt, and clay are known to influence the SWHC of any kind of soil, and they are commonly used in pedotransfer functions or PTF (Gupta, Dowdy & Larson 1977; Weil & Brady 2017). We included topographic variables as indirect predictors because these parameters have been shown to increase the quality of PTF (Sharma, Mohanty & Zhu 2006; Blanco *et al.* 2018). The elevation, aspect, and convexity were extracted from maps at 5 m resolution. Convexity being sensitive to scale, it was calculated with two different radii of moving windows, i.e., at 3 and 10 pixels. The selected variables presented a Spearman correlation coefficient < 0.70 (Dormann *et al.* 2012).

To overpass the risk of model overfitting due to the low number of SWHC measurements available (n=24) and the use of 8 predictors, we used a new approach *ensemble of small models* (Breiner *et al.* 2015; Breiner *et al.* 2017) specifically developed recently for modelling small sample sizes. We fitted bivariate models using the 28 possible combinations of pairs of

predictors with linear models (LM). Each bivariate model was evaluated by measuring an adjusted-R^2 through a leave-one-out cross-validation (form of jackknife). We selected only bivariate models with an adjusted-$R^2 > 0.5$ in the final ensemble and weighted them in the ensemble models based on the cross-validated adjusted-R^2 (Fig. 3). We then calculated an adjusted-R^2 between the value predicted by the final ensemble model and the observed values (Fig. 3). Last, we projected these models on the 256 plots distributed across the entire VA area (Fig. 2).

Measurements of Soil Water Available for Plants

We calculated three indices of SWAP.

1) **[pF2.7 – pF4.2]:** the theoretical SWAP found in the literature, corresponding to the difference between field capacity (pF=2.7) and wilting point (pF=4.2) (Gobat *et al*. 2004);

2) **[pF1.8 – pF4.2]:** a larger range of SWAP that also includes gravitational water;

3) **[pF2.7 – pF3.7]:** a more restricted SWAP range with a temporary wilting point (pF=3.7) that does not take into account the slowly availably water (Lal & Greenland 1979, Weil & Brady 2017).
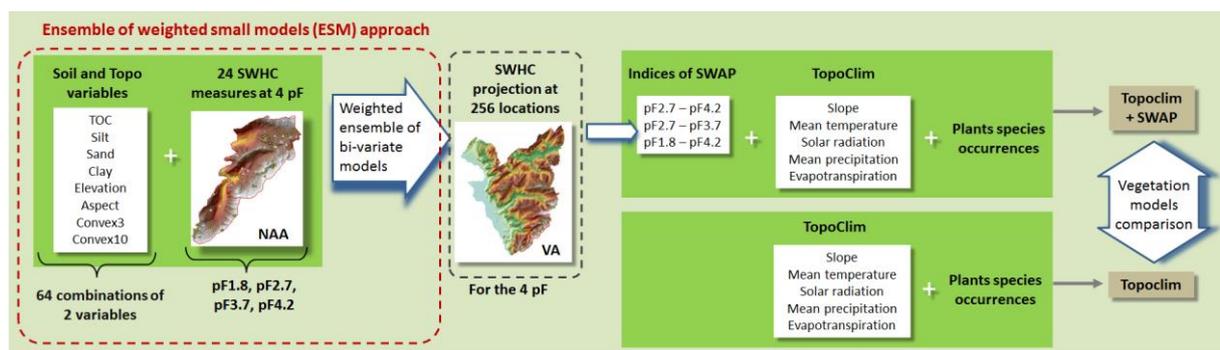

Fig. 3 – Flowchart of methods. SWHC: soil water holding capacity; SWAP: soil water available for plants.

Step 2 - Plant species distribution modelling

We modelled the distribution of 104 plant species (only herbaceous species were considered) over the 256 plots covering the VA area (Fig. 2). We considered only species with a minimum of 30 occurrences throughout the 256 plots (Buri *et al*. 2017). We ran models that consider four combinations of predictors. The first set of predictors – hereafter, *TopoClim* - was common to all models and included the following predictors, which were shown to be important for predicting plant species distributions (SDMs) (Randin *et al.* 2009a; Dubuis *et al.* 2013; Buri *et al.* 2017): the slope, mean temperature, solar radiation, mean precipitation and evapotranspiration (https://www.swisstopo.admin.ch/). The other three combinations of predictors included *TopoClim* and one of the different indices of SWAP: *TopoClim+[pF2.7-pF4.2], TopoClim+[pF2.7-3.7],* and *TopoClim+[pF1.8-pF4.2].* Finally, we randomized the values of the SWAP that gave the best evaluation values between the existing plots to create a new randomized SWAP variable. Thus, we maintained realistic ranges of measurements for SWAP but broke the link of each measurement with its geographical space. This randomized SWAP variable was used to run a fifth set of predictors: *TopoClim+pF-random.*

Plant SDMs were fitted in the R-3.5.0 statistical software using the biomod2 packages (Thuiller *et al.* 2016). We used a regression random forest to calibrate the model. The "randomForest" package of the R software, which implements the Breiman's algorithm, was used as well as the default parametrization of the function (i.e number of variables sampled randomly and proposed at candidate at each split = p/3, where p is the number of variables; number of grown trees = 500, trees are allowed to grow to the maximal number of terminal nodes). The selection of 80% of the data for model calibration and 20 % for model validation is a standard procedure used in many modelling papers (Buri *et al.* 2017; Guisan, Thuiller & Zimmermann 2017). We split the dataset into two partitions; 80% of the plots was used for model calibration, and the remaining 20% was used for independent evaluation. The predictive power of the models was determined by the maximization approach of the true skills statistics (TSS, Allouche, Tsoar & Kadmon

2006; for its maximization, maxTSS, see Guisan Thuiller & Zimmermann 2017), based on 100 repeated runs, which looks for the maximum value of TSS measured across a series of thresholds to binarize the probability predictions (as described in Guisan *et al.* 2017). TSS compares the number of correct predictions (minus those attributable to random guessing) to that of a hypothetical set of perfect predictions (Allouche, Tsoar & Kadmon 2006; Fernandes, Scherrer & Guisan 2019). TSS takes into account both omission and commission errors and is successful as a result of random guessing (TSS = Sensitivity + Specificity − 1). It ranges from − 1 to +1, where +1 indicates perfect agreement and values of zero or less indicate a performance not better than a random effect. This evaluation method has been proved recently by Fernandes, Scherrer and Guisan (2019) to better evaluate the fitness of models than the more standard AUC measure (Swets 1988). For each species, we also calculated the importance of each variable in the different models using the approach described below.

Step 3 - Plant species distribution model comparisons

We compared the performance of the five sets of models built with the different combinations of predictors by comparing their maxTSS. We also calculated the percentage of plant species that experienced a maxTSS increase due to the addition of SWAP as a supplementary predictor. For each set of models, the variable importance for each variable used in the models was fitted in the R 3.0.1 statistical software using the biomod2 packages (Thuiller *et al.* 2016). The package calculates the variables' importance by shuffling a single variable of the given data and then making a model prediction with this 'shuffled' data set. Then, the package computes a Pearson's correlation coefficient between the reference predictions and the 'shuffled' prediction. The return score is 1 minus the correlation between the references and the 'shuffled' prediction. The higher the value, the more influence the variable has on the model. A value of 0 assumes no influence of that variable on the model (Thuiller *et al.* 2016). Finally, we

calculated the mean change in the maxTSS values of all plant species present per plot in order to obtain an estimation of the change that occurs due to the addition of SWAP as predictors in each community. For this, vegetation alliances (Delarze & Gonseth 2008) were used with the average of delta maxTSS values per plot to determine which vegetation communities were most improved. Vegetation alliances were then grouped into four classes according to their characteristics: meadow, pasture, alpine grassland and scree.

Results

Evaluation of the bivariate models predicting SWHC at different pFs

From the 28 models of bivariate combinations, 7 showed an adjusted-$R^2 > 0.5$ for at least one SWHC pF model (Table 1). For SWHC at pF=1.8, pF=2.7 and pF=4.2, only one combination of bivariate variables gave an adjusted-$R^2 > 0.5$, i.e., the pairs' TOC-elevation, clay-convex3 and elevation-convex10, respectively. In contrast, for pF=3.7, six combinations of bivariate models gave an adjusted-$R^2 > 0.5$. Four of the combinations contained each time the soil texture as a predictor, and the two others contained the elevation with either convexity3 or convexity10. The ensemble model for SWHC pF=3.7 had an adjusted-R^2 of 0.63.

Comparisons of the plant species distribution models

The four sets of models (*TopoClim* alone and with the three different indices of SWAP) obtained predictions with average values of maxTSS ranging from 0.661 to 0.676 (Fig. SI2 and Table SI2 in Supplementary Information).

Variables used in the bivariate model	Water content at pF1.8	Water content at pF2.7	Water content at pF3.7	Water content at pF4.2
TOC, elevation	**0.519**	0.459	0.119	0.036
Clay, convex3	0.435	**0.551**	**0.660**	0.112
Clay, convex10	0.312	0.283	**0.597**	0.151
Silt, convex3	0.193	0.313	**0.567**	-0.018
Sand, convex3	0.254	0.394	**0.542**	-0.035
Elevation, convex3	0.367	0.407	**0.550**	0.305
Elevation, convex10	0.179	0.094	**0.567**	**0.627**
ESM	0.519	0.551	0.646	0.627

Tab. 1 – Leave-one-out-validation values of bivariate models and ensemble of weighted-small-models (ESM) using an adjusted-R^2 for each water content modelled.

We found that adding indices of SWAP improved the overall model quality and that maxTSS especially increased significantly (paired t-test, p-value <0.05) when *[pF1.8-pF4.2]* was included (Fig. 4-a, Table SI4 in Supplementary Information). The greatest increase in maxTSS was found by adding *[pF1.8-pF4.2]* (Fig. 4-a, Tab. 2). For some species (i.e., *Hieracium lactucella, Poa trivialis, Alchemilla xanthochlora, Crocus albiflorus,* and *Trifolium thalii*), the addition of one of the SWAP variables increased the maxTSS of their model by more than 10% (Tab. 2). The highest increase of 40% was observed when *[pF1.8-pF4.2]* was added to the model of *Hieracium lactucella* (Table 2). Overall, the addition of the three different SWAP ranges induced a maxTSS increase from 45% to 56% for the plant species considered (Fig. 4-b). Specifically, the addition of *[pF1.8-pF4.2]* to *TopoClim* caused an increase greater than 10% of maxTSS for 14% of the plant species (Fig. 4-b).

In general, the minimum temperature was the most important variable in all four sets of models for the majority of species. When we added the various *SWAP* predictors, *[pF1.8-pF4.2]* became the third most important variable set for predicting plant species distribution, after temperature and precipitation (Fig. 5).
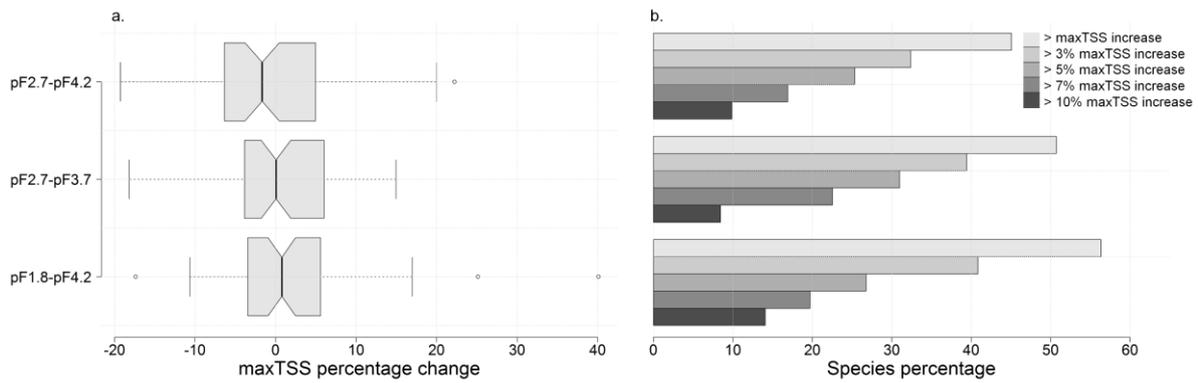
Fig. 4 – a. Boxplot of the maxTSS percentage change after adding different SWAP ranges. Outliers were defined in the boxplots as data points outside the fences ("whiskers") of the boxplot (outside refers to 1.5 times the interquartile range above the upper quartile and bellow the lower quartile). No outlier was removed from any analysis in our study, these outliers are only appearing visually in the boxplots. **b**. Bar plot of the percentage of species with improved maxTSS after adding one of the SWAP variables.
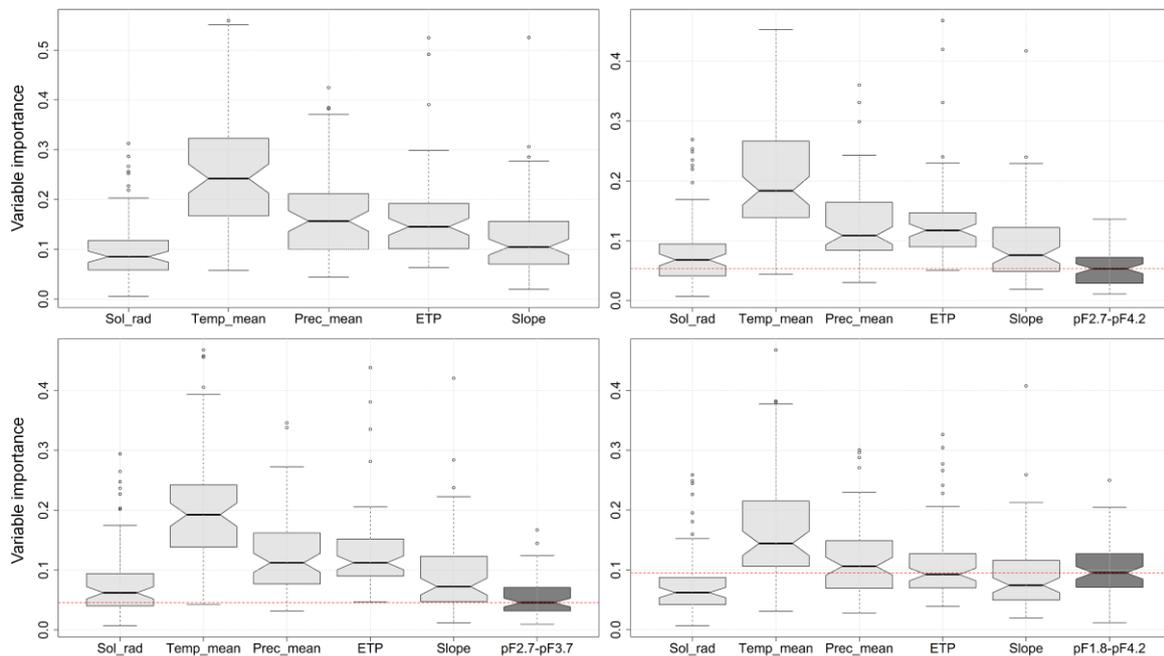

Fig. 5 – Boxplot of variable importance calculated or each set of models. Sol_rad: solar radiation, Temp_mean: mean temperature, Prec_mean: mean precipitation, ETP: evapotranspiration, Slope: slope at the site. Outliers were defined in the boxplots as data points outside the fences ("whiskers") of the boxplot (outside refers to 1.5 times the interquartile range

above the upper quartile and bellow the lower quartile). No outlier was removed from any analysis in our study, these outliers are only appearing visually in the boxplots.

Looking at plant communities, we found that, in general, meadows and pastures were the most improved by adding SWAP *[pF1.8-pF4.2]* as a predictor (Fig. 4). Overall, meadows showed the best improvements, while alpine grasslands in rocky conditions benefitted the least from the addition of the different SWAP indices, i.e., *Carex* dry calcareous grasslands (*Caricion firmae*), *Sesleria* dry calcareous grasslands (*Seslerion*), and calcareous snow beds (*Arabidion caeruleae*). Among the screes, one of the plots belonging to the wet calcareous screes (*Petasition paradoxi*) showed the highest maxTSS increase by adding SWAP *[pF1.8-pF4.2]* (Fig. 6).
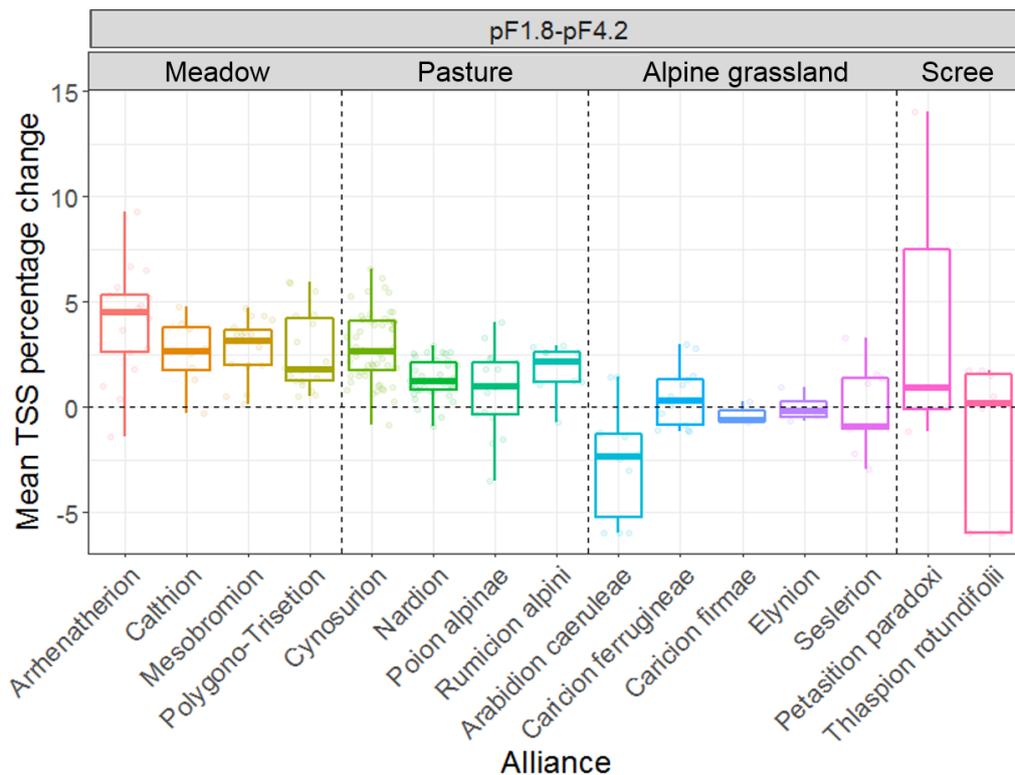

Fig. 6 - Boxplot of maxTSS percentage change after adding the different SWAP ranges in relation to the alliances that were determined to be the average of the delta maxTSS values per plot.

Discussion

In this study, we showed that using the weighted ensemble of small models (ESM) approach allowed predicting the SWHC over a large and complex mountain area, from models calibrated in a smaller subarea. In addition, it has been emphasized that including different predicted ranges of SWAP derived from SWHC as predictors in plant species distribution models (plant SDMS) could significantly improve the SDM predictive capacity for many mountain plant species. Moreover, when added to the set of basic topo-climatic predictors, SWAP *[pF1.8-pF4.2]* allowed improving model performance.

Ensemble of Small Models

Only a few studies attempted to spatialize SWHC and used it subsequently in models of plant distributions (plant SDMs). For instance, Piedallu *et al.* (2011) investigated how the use of soil information from numerous forest plots achieved a fine-resolution SWHC map over a large area. To do this, they evaluated six classes of pedotransfer functions to convert the soil texture in SWHC; then, they used the results of the best pedotransfer functions to map the SWHC using ordinary kriging, which is a geostatistical method that is very different from the predictive approach used in this study. The authors employed a large set of data (> 100,000 sampled plots) to build their pedotransfer functions and to interpolate the obtained SWHC. This is an isolated example of a study where a very large dataset was available, but the majority of studies have to face a scarcity of SWHC measurements (and, in general, a scarcity of soil property measurements). In the latter cases, our methodology represents a pertinent solution to generalize SWHC spatially from only a few measurements. Moreover, it includes more environmental variables without the risk of overfitting the models (Breiner *et al.* 2015; Breiner *et al.* 2017; Di Cola *et al.* 2017.).

It is known that SWHC depends on organic matter content and soil texture (Weil & Brady 2017). Organic matter content plays a role on the support of the soil structure and porosity and increases the water holding capacity of the soil (Brady & Weil 2010). Organic matter interacts with soil texture and, globally, the finer the grain size, the more water is retained by the soil. Thus, a fine texture avoids losing too much water through percolation due to gravity (Gobat, Aragno & Matthey 2004). Texture and organic matter are taken into account in the pedotransfer functions to obtain SWHC values (Bastet *et al.* 2008). Nevertheless, as SWHC is expected to be variably influenced by these two variables in different soils (Brady *et al*. 2008), the use of topographic predictors can help to refine their incorporation into the models or can be used as proxies for other factors that explain SWHC variation (e.g., soil depth, type of pedogenesis) that cannot be measured across large areas. Some studies even used only one predictor derived from the Landsat images and a digital elevation model (Blanco *et al.* 2018) to spatialize the soil water retention capacity. In our study, topographic variables were significantly selected as predictors for all pF values.

Organic matter content was able to correctly capture the spatial variation of SWHC at pF=1.8, and the texture was useful for pF=2.7 and 3.7. Conversely, these variables were not able to reflect the variation of pF=4.2, for which only topographic variables were selected. This result suggests that organic matter is able to retain soil water when water is abundant but that when water becomes rare, texture and topography (as a surrogate of soil depth) have important roles; for that reason, topographic information is more useful for modelling soil response to dry conditions.

Improvement of plant SDMs with SWAP

In general, plant SDMs were significantly improved by the addition of SWAP calculated from the different SWHCs as a predictor. A greater improvement was given by SWAP *[pF1.8-*

pF4.2], which includes the largest ranges between the wilting point and the field capacity tested in our study. Hence, the better predictive power of SWAP *[pF1.8- pF4.2]*, compared to that of the other range values, may correspond to the fact that this larger estimation of available water can discriminate better the establishment of plant populations. We can hypothesize that with lower pF values, water becomes more available to all plant species, but pF=1.8 appeared to be more discriminant than the pF classically used in the literature (i.e., pF=2.7) to define the amount of water available for plants.

We highlighted that the available water started to be discriminant among species when the pF values exceeded 4.2. This finding might translate to responses that are modulated by various plant adaptations: sites with high SWAP (above these values) can probably be colonized easily by most species, with an advantage then given to the most competitive species in productive conditions (Grime's C strategy; (Grime 2006)), whereas meso- or xerophilous species tend to find advantageous conditions in drier soils with lower SWAP values. In fact, most of the species showing the strongest improvement in adding SWAP have a C strategy, or more precisely, species living in deep soils (Table 2). It is important to note that plant species that have long roots could overpass the limitation of a low water holding capacity in the upper soil by reaching water in deeper parts of the soil. Thus, those species would not be influenced by the topsoil SWAP, as measured in this work.

Looking at habitat types, plants inhabiting deeper soils and lowlands (e.g., *Arrhenatherion, Mesobromion, Calthion,* and *Polygono-Trisetion*) were better predicted by the addition of the SWAP in their SDMs, and conversely, alpine grasslands in rocky conditions (*Caricion firmae, Seslerion,* and *Arabidion caeruleae*) showed poor improvements, if any.

Species	TSS percentage change	
Poa trivialis	22.25	
Hieracium lactucella	19.99	
Alchemilla xanthochlora	18.05	
Heracleum sphondylium	17.46	
Chaerophyllum hirsutum	13.96	pF2.7 – pF4.2
Crocus albiflorus	10.66	
Geranium sylvaticum	10.27	
Trifolium thalii	9.89	
Ranunculus nemorosus	9.87	
Galium anisophyllon	8.5	
Heracleum sphondylium	14.97	
Hieracium lactucella	14.84	
Deschampsia cespitosa	11.99	
Ajuga reptans	11.99	
Poa trivialis	11.79	pF2.7 – pF3.7
Chaerophyllum hirsutum	11.49	
Trifolium thalii	9.57	
Trifolium pratense	9.57	
Crocus albiflorus	9.55	
Briza media	9.2	
Hieracium lactucella	40.11	
Alchemilla xanthochlora	25.15	
Crocus albiflorus	16.98	
Carum carvi	16.66	
Hypericum maculatum	14.42	pF1.8 – pF4.2
Carex flacca	14.06	
Anthoxanthum odoratum	12.28	
Trifolium thalii	11.34	
Trifolium pratense	10.45	
Poa trivialis	10.01	

Tab. 2 – List of the 10 species that had the greatest improvement of their maxTSS by adding one of the SWAP variables.

In their review of which predictors are necessary, available and used to model plant species distributions, Mod *et al.* (2016) found only 15 out of 200 studies that incorporated SWAP or soil moisture in SDMs, and only three of these focused on mountain regions. Mellert *et al.*

(2011) showed the usefulness of adding available water capacity even at a coarse resolution (1:200,000) to model tree species distributions in the Bavarian Alps. Boulangeat, Gravel and Thuiller (2012) also incorporated a coarse estimate of soil water capacity extracted from a 1 km resolution European soil database to model plant species distributions. Gallien *et al.* (2012) did the same with a water holding capacity that originated from the Aurelhy meteorological model (Bénichou & Le Breton 1987). In this study, we observed the same trend in the improvement of plant distribution models at a much higher resolution and with precise SWHC measurements to calculate the SWAP indices. We could infer that SWAP influences the vegetation at both a small and large scale and that some SWAP ranges (e.g., *[pF1.8-pF4.2]*) have a greater predictive power than others.

SWAP can convey into the plant models both direct and indirect information about water availability and its effect on plant growth and survival, which would be lacking in models based on topo-climatic variables only. First, soil water availability has a direct effect on plant germination and establishment and, thus, on plant distribution. It has been shown that drought is one of the main seedling mortality causes in alpine environments (Venn & Morgan 2009). However, Körner and Life (2003) showed that a low soil water content may limit nutrient cycling and bacterial activity, which could in turn reduce nutrient supply to plants. In that sense, SWAP plays an indirect role on plant species distributions through nutrient availability, as the water content not only depends on soil texture and OM proportions, such as nutrients, but it also governs the exchanges between plants and nutrients through the soil solution quality (Doran & Parkin 1999).

Soil moisture measured in the field has also been proved to be an important predictor in plant SDMs, but such direct measurements of soil wetness have a large temporal variability (le Roux, Aalto & Luoto 2013). In contrast, SWAP is mainly determined by soil texture and, thus, is more stable throughout time (Gobat, Aragno & Matthey 2004) and, therefore, likely better represents

long-term plant species distribution—i.e., the soil water supply relationship. We showed the importance of combining SWAP values with precipitations (which represent the total water entering the system) in plant species distribution models. These indices with a physiological significance could be used to characterize water availability for plants over large areas, thereby improving plant distribution models and their potential productivity or diversity. This knowledge is particularly important in the climate change context to identify and monitor potential impacts on vegetation.

Conclusions, limitation, and perspectives

SWHC is a quite complex analysis to carry out, especially costly and time consuming, which explains the low sample size of SWHC measurements. It therefore also explains why this variable is still very rarely included in spatial modelling studies of plant distributions. An important side aspect of this study was to use a newly developed approach *ensemble of small models* (see Breiner *et al.* (2015); Breiner *et al.* (2017) here using an ensemble of bivariate models) that can deal with very low sample sizes for the response variable (here SWHC modelled from 24 locations) while still accounting for multiple predictors in the model and allowing spatial predictions. In this regard, we feel that our approach remains a very valuable first step in the right direction, by proposing such methodology to 'spatialize' variables that are difficult to measure and thus have usually low sample size. The proposed ESM approach could therefore also be applied to other complex soil variables for which few data are available.

We used a weighted ensemble of small bivariate models to circumvent the small number of pF values that were measured in the field and prevented the inclusion of more than two variables at a time in a single model. Because we already obtained pertinent results, it is likely that with a higher sample size, we would probably have obtained better predictions. However, a trade-off between costly field measurements and gains in model quality should be investigated to

obtain the minimum number of additional field samples required to obtain reliable models. In our case, we analysed four different pF variables to assess which one could bring more information to our models. For future studies, we would suggest focusing on the few most predictive pFs identified in this study, such as pF=1.8 and pF=4.2, and collect many more samples in the geographic space to increase model quality. Moreover, sampling shallow soils with too high lithological content proved very difficult or even impossible in some cases. Thus, as these types of soil were *de facto* excluded from the analysis, we must be careful when applying our model to areas that are characterized by such types of soils. This consideration could explain the high variability in the model outputs of scree habitat types, for instance.

In this study we only considered the 15 cm top soil, while being aware that many plants can extract water from much deeper than 15 cm in the soil profile, especially at lower elevations. However, in our case of subalpine/alpine high-elevation plant communities, and although the description of soil profiles at each of the 24 sampling locations indicated rooting depth in the range 0 to 60 cm , the highest root density was always observed in the 0-15 cm layer (Zingg 2015). In profiles with deep rooting, there was no sharp textural discontinuity between the 0-15 cm and 15+ cm layers, the texture being one of the main determinant of SWHC. Therefore, if texture is homogeneous through the soil profile, we expected that SWHC should not be completely different near the surface and at greater depth. As a result, we choose to use 15 cm soil cores as an indicator of water holding conditions at each site. Ideally, one should consider for each plant species a soil sample collected at different depths depending on the length of its root network. This would however be very difficult in practice and time consuming, also reducing our ability to make spatial predictions. Moreover, the upper soil is the most responsive part to fluctuations of water availability, the first to dry and the first to be wet after short rains. It also plays an essential role in plant nutrition, with nutrient availability strongly declining when dry. For all these reasons, our soil samples confined to the upper 15 cm of soil should

remain a reasonable proxy of available water and a good tradeoff between the time spent in the lab to perform analyses and the ability of a measurement to capture information for both shallow rooting species and deeper rooting plant species.

In this study we used topographic variables and soil properties as predictors of different pF variables, which represent soil moisture content at different standardized pressures in laboratory conditions and therefore serve as estimates of the SWHC. It is expected that SWHC is a direct representation of soil porosity parameters (mainly soil organic matter and soil texture), which are then used to fit pedotransfer functions. We supposed that topographic variables could indirectly predict SWHC values. Our study interestingly revealed that pF=4.2 is indeed more influenced by topographic features than by organic matter or texture measured in the field. However, being based on a correlative approach and a limited number of samples, our results cannot establish a formal causal link between topography and SWHC. We can presume that topographic features act as a surrogate of texture at a larger scale that considers neighbourhood features rather than a precise point data. Combining pF differences with precipitations allowed for the estimation of soil water availability for plants, but similarly, with such a correlative approach, we cannot establish a formal causal link between SWAP and plants species distributions. Complementary field experiments, for instance, those deployed along a gradient of SWHC or SWAP predictions such as in our models, would be necessary to establish such a causal link between either the topography and SWHC or between the SWAP and plants distributions. An explanation we can propose for the link between SWHC and topography observed here is the strong glacial history of the region (Joerin, Stocker & Schlüchter 2006). Moraines and loess have been deposited everywhere during the glacial periods. Wind erosion then displaced the fine elements, which are important for water retention, from sloping and convex surfaces to concave surfaces.

We also showed that SWHC models and derived SWAP indices combined with precipitation improved the predictions of grassland plant species distributions in the Western Swiss Alps, across a large range of elevations. It would be of interest to repeat this study in other regions, using other topographic, climate, and soil conditions, to test whether similar results can be obtained and to what extent the SWHC models and SWAP indices can be transferred to other areas and different environments.

Acknowledgement

References

Adhikari, K. & Hartemink, A.E. (2016) Linking soils to ecosystem services - A global review. *Geoderma,* **262,** 101-111.

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

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.

Bastet, G., Bruand, A., Quétin, P. & Cousin, I. (2008) Estimation des propriétés de rétention en eau des sols à l'aide de fonctions de pédotransfert (FPT) : une analyse bibliographique. . *Etude et Gestion des Sols,* **5**.

Bénichou, P. & Le Breton, O. (1987) Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *La Météorologie,* **7**.

Blanco, C.M.G., Gomez, V.M.B., Crespo, P. & Liess, M. (2018) Spatial prediction of soil water retention in a Paramo landscape: Methodological insight into machine learning using random forest. *Geoderma,* **316,** 100-114.

Boulangeat, I., Gravel, D. & Thuiller, W. (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters,* **15,** 584-593.

Brady, N.C. & Weil, R.R. (2010) Elements of the nature and properties of soils. Pearson Educational International Upper Saddle River, NJ.

Breiner, F.T., Guisan, A., Bergamini, A. & Nobis, M.P. (2015) Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution,* **6,** 1210-1218.

Breiner, F.T., Guisan, A., Nobis, M.P. & Bergamini, A. (2017) Including environmental niche information to improve IUCN Red List assessments. *Diversity and Distributions,* **23,** 484-495.

Buckman, H.O. & Brady, N.C. (1922) *The nature and properties of soils*. Macmillan.

Buri, A., Cianfrani, C., Pinto-Figueroa, E., Yashiro, E., Spangenberg, J.E., Adatte, T., Verrecchia, E., Guisan, A. & Pradervand, J.-N. (2017) Soil factors improve predictions of plant species distribution in a mountain environment. *Progress in Physical Geography,* **41,** 703-722.

Calanca, P. (2007) Climate change and drought occurrence in the Alpine region: How severe are becoming the extremes? *Global and Planetary Change,* **57,** 151-160.

De Vos, B., Van Meirvenne, M., Quataert, P., Deckers, J. & Muys, B. (2005) Predictive quality of pedotransfer functions for estimating bulk density of forest soils. *Soil Science Society of America Journal,* **69,** 500-510.

Delarze, R. & Gonseth, Y. (2008) Guide des milieux naturels de Suisse. *Rossolis*.

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F.T., D'amen, M., Randin, C., Engler, R., Pottier, J., Pio, D. & Dubuis, A. (2017) ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography,* **40,** 774-787.

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C. & Schröder, B. (2012) Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography,* **39,** 2119-2131.

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.

Fernandes, R.F., Scherrer, D. & Guisan, A. (2019) Effects of simulated observation errors on the performance of species distribution models. *Diversity and Distributions*.

Gallien, L., Douzet, R., Pratte, S., Zimmermann, N.E. & Thuiller, W. (2012) Invasive species distribution models - how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography,* **21,** 1126-1136.

Gobat, J.-M., Aragno, M. & Matthey, W. (2004) *The living soil: fundamentals of soil science and soil biology*. Science Publishers.

Grime, J.P. (2006) *Plant strategies, vegetation processes, and ecosystem properties*. John Wiley & Sons.

Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017) *Habitat suitability and distribution models: with applications in R*. Cambridge University Press.

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

Gupta, S.C., Dowdy, R.H. & Larson, W.E. (1977) Hydraulic and Thermal-Properties of a Sandy Soil as Influenced by Incorporation of Sewage Sludge. *Soil Science Society of America Journal,* **41,** 601-605.

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

Jasper, K., Calanca, P. & Fuhrer, J. (2006) Changes in summertime soil water patterns in complex terrain due to climatic change. *Journal of Hydrology,* **327,** 550-563.

Kammer, P.M., Schob, C., Eberhard, G., Gallina, R., Meyer, R. & Tschanz, C. (2013) The relationship between soil water storage capacity and plant species diversity in high alpine vegetation. *Plant Ecology & Diversity,* **6,** 457-466.

Körner, C. & Life, A.P. (2003) Functional plant ecology of high mountain ecosystems. *Alpin Plant Life. Berlin; Heidelberg*, pp. 345.

Lafargue, E., Marquis, F. & Pillot, D. (1998) Rock-Eval 6 applications in hydrocarbon exploration, production, and soil contamination studies. *Revue De L Institut Francais Du Petrole,* **53,** 421-437.

le Roux, P.C., Aalto, J. & Luoto, M. (2013) Soil moisture's underestimated role in climate change impact modelling in low-energy systems. *Global Change Biology,* **19**.

Mellert, K.H., Fensterer, V., Kuchenhoff, H., Reger, B., Kolling, C., Klemmt, H.J. & Ewald, J. (2011) Hypothesis-driven species distribution models for tree species in the Bavarian Alps. *Journal of Vegetation Science,* **22,** 635-646.

Mod, H.K., Scherrer, D., Luoto, M. & Guisan, A. (2016) What we use is not what we know: environmental predictors in plant distribution models. *Journal of Vegetation Science,* **27,** 1308-1322.

Patil, N.G. & Singh, S.K. (2016) Pedotransfer Functions for Estimating Soil Hydraulic Properties: A Review. *Pedosphere,* **26,** 417-430.

Piedallu, C., Gegout, J.C., Bruand, A. & Seynave, I. (2011) Mapping soil water holding capacity over large areas to predict potential production of forest stands. *Geoderma,* **160,** 355-366.

Pribyl, D.W. (2010) A critical review of the conventional SOC to SOM conversion factor. *Geoderma,* **156,** 75-83.

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

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

Richards, L. & Weaver, L. (1944) Moisture retention by some irrigated soils as related to soil moisture tension. *Journal of Agricultural Research,* **69,** 215-235.

Sharma, S.K., Mohanty, B.P. & Zhu, J.T. (2006) Including topography and vegetation attributes for developing pedotransfer functions. *Soil Science Society of America Journal,* **70,** 1430-1440.

Shirazi, M.A. & Boersma, L. (1984) A Unifying Quantitative-Analysis of Soil Texture. *Soil Science Society of America Journal,* **48,** 142-147.

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

Thuiller, W., Georges, D., Engler, R. & Breiner, F. (2016) biomod2: ensemble platform for species distribution modeling. R Core Team.

Veihmeyer, F. & Hendrickson, A. (1928) Soil moisture at permanent wilting of plants. *Plant Physiology,* **3,** 355.

Venn, S.E. & Morgan, J.W. (2009) Patterns in alpine seedling emergence and establishment across a stress gradient of mountain summits in south-eastern Australia. *Plant Ecology & Diversity,* **2,** 5-16.

Viji, R. & Rajesh, P.P. (2012) Assessment of water holding capacity of major soil series of Lalgudi, Trichy, India. *J. Environ. Res. Develop,* **7**.

Weil, R.R. & Brady, N.C. (2017) The nature and properties of soils, 15th. *Pearson***,** 621-624.

Wosten, J.H.M., Pachepsky, Y.A. & Rawls, W.J. (2001) Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology,* **251,** 123-150.

Zingg, B. (2015) Modélisation de la reserve hydrique des sols dans les Alpes vaudoises méridionales. Université de Lausanne.
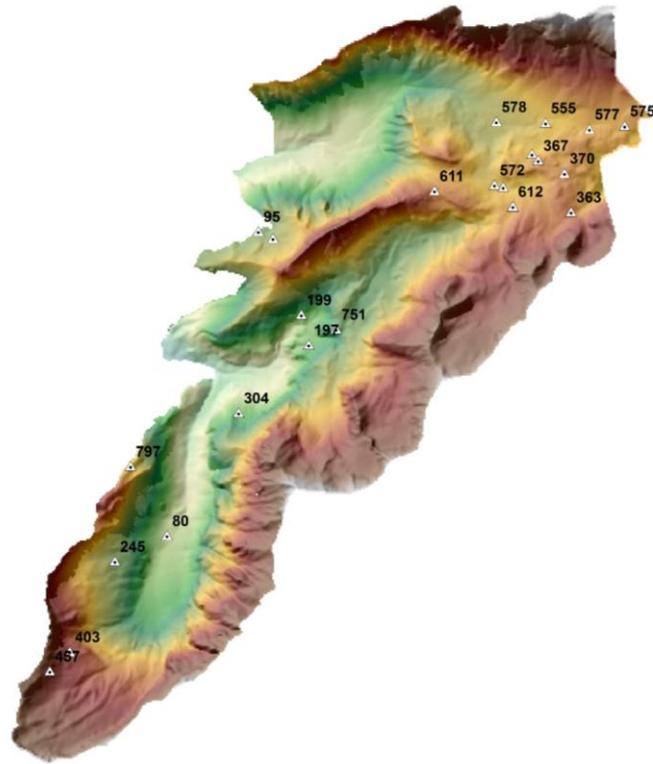
Supplementary Information

Fig. SI1 – Study area and location of plots.

Plot	OM [%]	TOC [%]	Clay [%]	Silt [%]	Sand [%]	pF 1.8 [cm³/cm³]	pF 2.7 [cm³/cm³]	pF 3.7 [cm³/cm³]	pF 4.2 [cm³/cm³]	Soil type according to the World Reference Base (IUSS Working Group WRB, 2015)
80	4.8	2.8	8.3	66.8	25.0	28.6	25.5	20.9	6.9	Dystric Cambisol
95	22.0	12.8	9.1	68.9	22.0	33.7	30.0	28.2	12.8	Dystric Cambisol
141	10.4	6.0	10.3	59.1	30.5	33.2	28.6	27.0	13.7	Dystric Cambisol
197	20.8	12.0	11.6	73.2	15.2	27.6	25.6	24.4	10.9	Dystric Cambisol
199	17.9	10.4	10.9	63.3	25.8	24.6	21.9	20.4	10.7	Eutric Leptosol
245	10.5	6.1	10.4	71.0	18.6	33.9	29.7	27.7	17.4	Dystric Cambisol
304	11.7	6.8	13.3	54.0	32.7	27.1	24.1	22.6	11.1	Eutric Cambisol
363	68.4	39.7	13.4	70.5	16.1	18.6	16.8	16.0	7.3	Hyperdystric Cambisol
367	33.6	19.5	17.9	65.7	16.4	24.2	22.1	20.8	13.3	Eutric Cambisol
370	12.8	7.4	15.1	65.7	19.2	30.0	26.6	24.5	14.6	Hypereutric Cambisol
403	1.8	1.0	2.9	21.8	75.3	35.2	32.4	28.7	10.9	Orsteinic Podzol
457	43.1	25.0	12.2	54.4	33.3	23.2	20.5	19.1	9.2	Luvic Cambisol
555	25.5	14.8	21.8	52.9	25.4	29.6	26.4	23.8	15.3	Dystric Cambisol

565	28.4	16.5	18.9	71.8	9.3	28.0	25.1	NA	17.8	Stagnic Dystric Cambisol
572	41.8	24.2	14.7	68.9	16.4	17.4	15.9	NA	11.5	Endogleyic Stagnosol
575	16.6	9.6	21.1	65.6	13.3	28.3	25.3	23.2	15.4	Hypereutric Cambisol
577	18.8	10.9	17.8	71.7	10.5	35.9	32.7	29.8	23.6	Stagnosol
578	5.8	3.3	21.7	67.3	11.0	34.3	31.2	27.2	19.8	Dystric Cambisol
609	30.4	17.6	12.4	64.9	22.7	22.7	21.0	20.3	NA	Dystric Cambisol
611	4.1	2.4	7.3	434.0	48.7	31.2	26.1	19.4	6.4	Hyperdystric Cambisol
612	19.0	11.0	10.4	39.9	49.7	30.4	27.4	24.4	14.2	Dystric Cambisol
751	13.3	7.7	16.1	56.6	27.3	29.5	26.3	24.3	16.6	Hypereutric Cambisol
788	9.8	5.7	19.1	72.0	8.8	33.2	29.8	NA	23.2	Hypereutric Cambisol
797	15.6	9.1	14.3	48.2	37.5	26.0	22.4	NA	17.5	Dystric Cambisol

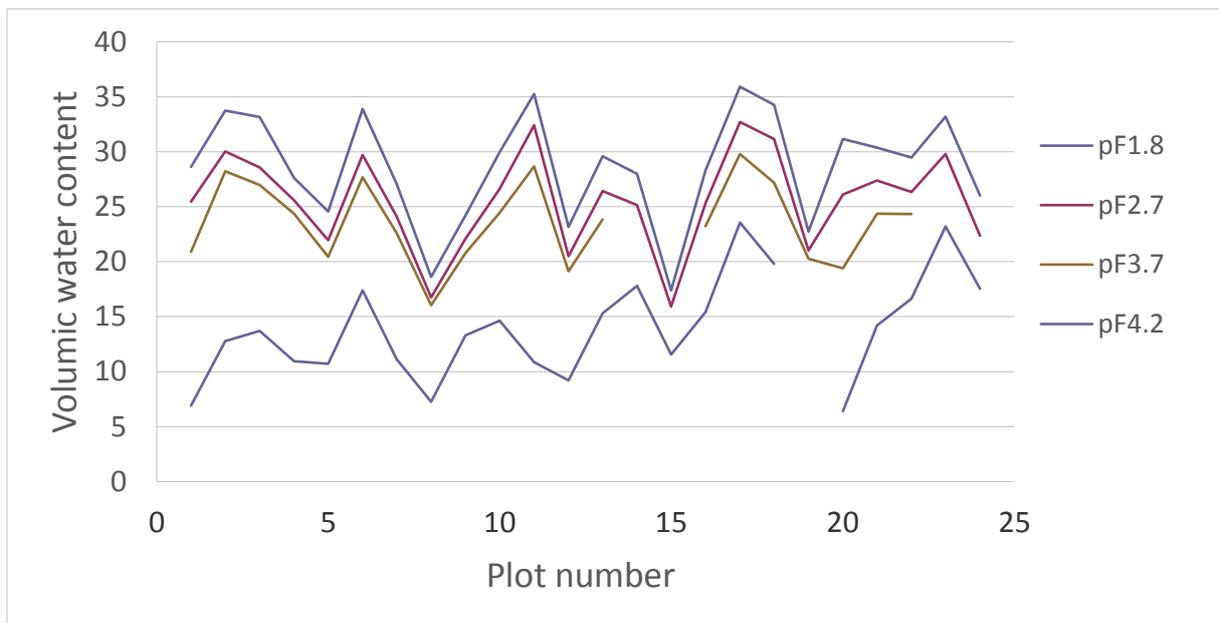**Tab. SI 1** – Soil properties measured

Fig. SI2 – Soil water holding capacity values measured at different pF. For plots n° 14, 15, 23 no measurements were available for pF3.7.

Table SI2 – Means and standard deviations of Total Organic Carbon (TOC) content, and sand, silt, and clay proportions in the NAA sector (24 plots) and VA sector (256 plots).

	TOC	clay	silt	sand
Mean NAA sector	11.77	13.7968189	60.76	25.44
Sd NAA sector	8.77	4.84376544	12.51	15.44

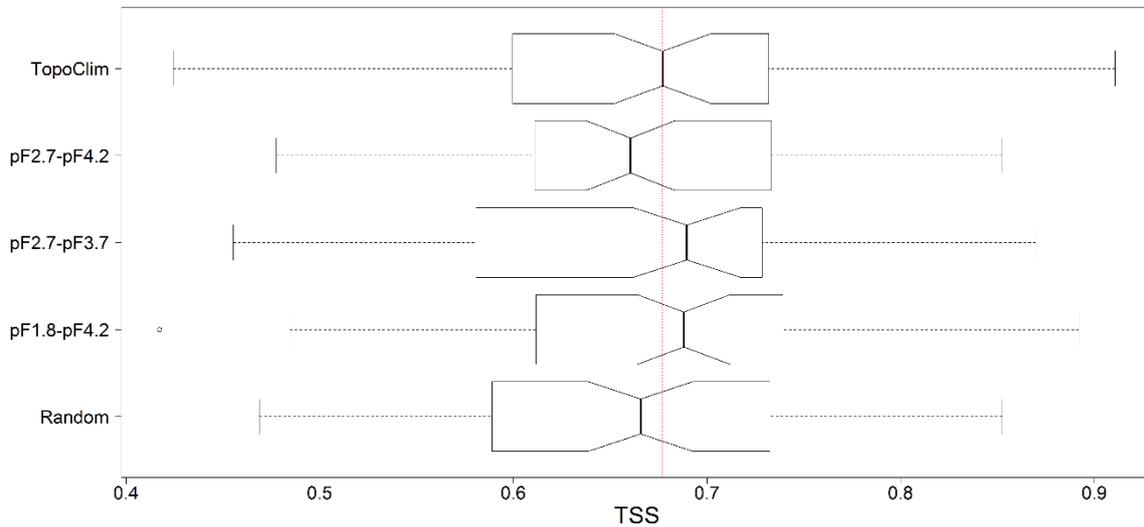|				
Mean VA sector	22.36	20.7924357	57.49	21.71
SD VA sector	11.73	5.83003459	6.65	9.41

Fig. SI3 – Boxplot of TSS values for each set of models.

Table SI3 – Evaluation of plants species distribution models for the five sets of predictors through TSS (mean).

Model	Mean TSS
TopoClim+[pF1.8-pF4.2]	0.676
TopoClim+[pF2.7-3.7]	0.667
TopoClim+[pF2.7-pF4.2]	0.661
TopoClim+pF-random	0.662
TopoClim	0.665

Table SI4 – Paired t-test between the *Topoclim* set of model and the *TopoClim+one of the 3 different ranges of SWAP*. The t-statistic, df and p-value of each test are shown.

	t-statistic	df	p-value
TopoClim+[pF1.8-pF4.2]	-1.8623	70	0.06675
TopoClim+[pF2.7-3.7]	-0.39578	70	0.6935
TopoClim+[pF2.7-pF4.2]	0.70965	70	0.4803