

# Introducing process-based disturbance variables in predictive models of plant distribution



Diploma thesis (university of Lausanne, 2005)

Grégoire Vuissoz

Under the supervision of the Professor Antoine Guisan

## Abstract

Alpine environment with its great landscape variability is a challenge for modeling the plant species distribution. The emergence of powerful statistical methods (e.g. GLM and GAM) as well as geographic information system (GIS) allows the prediction of potential habitat of plant species using spatial observation (presence/absence or abundance data) and environmental factors. However at small scale, the topo-climatic variables usually used become insufficient for accurate predicting. Searching for new predictors able to explain the part of deviance leftover by topo-climatic predictors is a way to improve the statistic models. In this study we integrate process-based disturbance variables like geomorphologic perturbation index and snow redistribution pattern in static models to see how they can improve predictive power for sub-alpine and alpine plant species. Results show that process-based disturbance variables are able to increase the predictive capacity of the models for about 15% of the tested species.

## Introduction

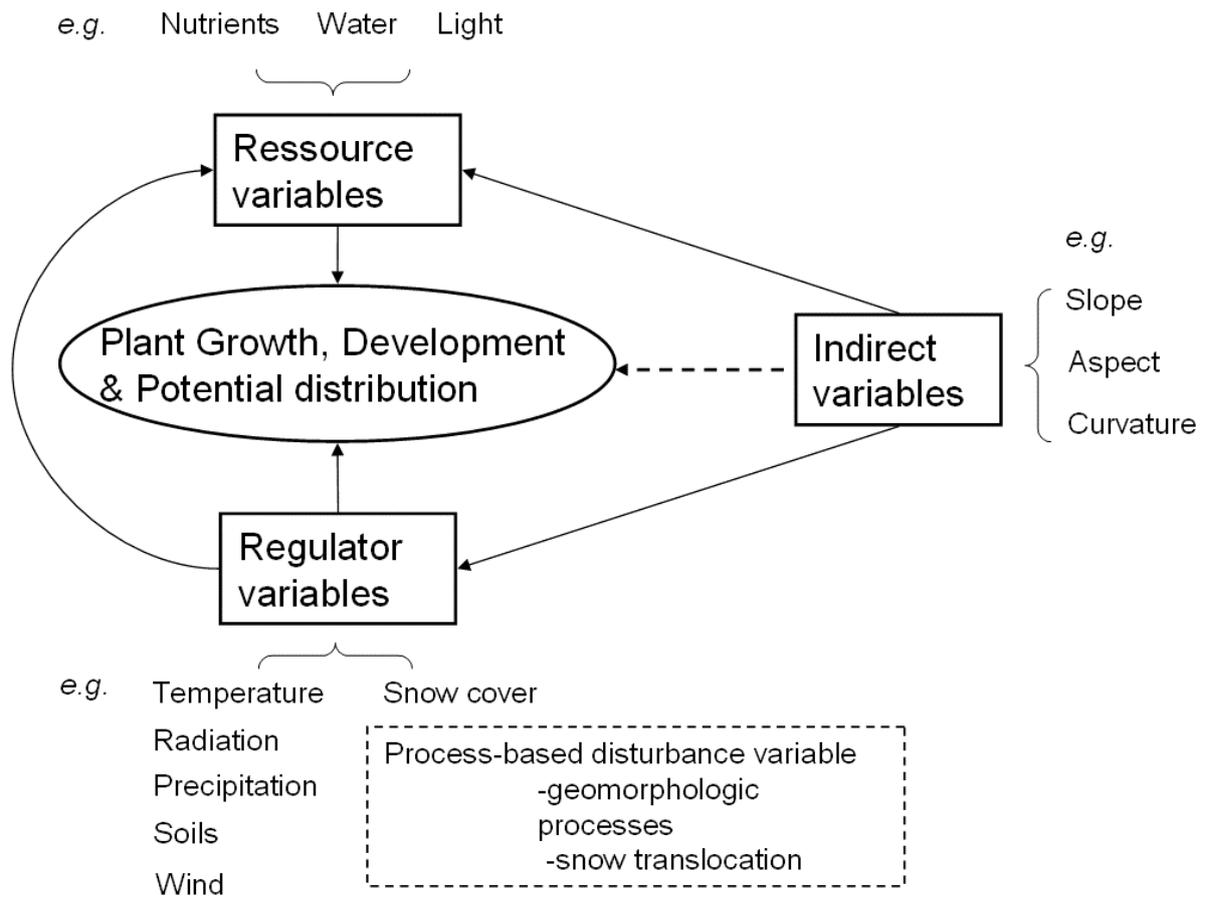
Alpine environment with its great landscape variability is a challenge for modeling the plant species distribution. Large altitudinal gradients and contrasted energy fluxes between exposed and shaded faces can explain the great observed biodiversity. Hence, it appears that this environment is one of the most sensitive of Europe's terrestrial ecosystems (Alpine convention, the international convention on the protection of the Alps). The monitoring of the alpine system through satellite remote sensing (ALPMON, Waser *et al.*, 2000) and the study of the vegetal species under changing climate conditions (ALPLANDI, Guisan 1999; EUROMOVE Bakkenes *et al.*, 2002) are great examples of what the scientific community can do in order to understand this ecosystem.

The emergence of powerful statistical methods (e.g. GLM and GAM) as well as geographic information system (GIS) allows the prediction of potential habitat of plant species using spatial observation (presence/absence or abundance data) and environmental factors. These models are useful for testing biogeographical hypothesis, for the management and the conservation of rare species or for the assessment of future climate change impact on plant species (Zimmerman and Kienast, 1999; Guisan *et al.*, 1998; Guisan and Theurillat, 2000). Static statistical models should be built with precision and precaution if we want to apply them for predicting future or past distributions (Thuiller, 2003). At large and meso-scale, like at the Switzerland scale, models based on the topo-climatic niche of the target species provide very good results. However the more we go down in scale, the more topo-climatic variables become insufficient for accurate predicting (Guisan *et al.*, 1998). Especially in regions having great landscape heterogeneity (like alpine environment), the realized niche of species is not only driven by the topo-climatic variables but also by spatial / geographical particularities. For instance, local stations separated from a few meters could experience drastically different conditions even if corresponding topo-climatic data are similar (Körner, 2003).

In order to avoid these pitfalls in plant distribution modeling we need to use predictors (*i.e.* predictive variables) that have a direct physiological or physical impact on plant species. But a current limitation of these models is the lack of spatially explicit predictors related to stresses or modifications occurring on the soil surface like anthropic-linked land use or hydro-mechanical processes.

The particular situation of alpine landscape leads to very active and dynamic hydro-mechanical processes like the geomorphologic processes described by the geographers (Rappaz & Hunziker, 1996; Loye & Pahud, 1996; Schoeneich *et al.*, 1998). These processes affect vegetation at the local scale (Balcerktewics S. *et al.*, 1985) but they modify also the species diversity at larger scale (Nichols *et al.* 1998). Gravity or hydrodynamic processes like rock slides or avalanche paths are examples of processes that have a big impact on the soil surface. Those essential components of the alpine environment strongly drive the species distribution patterns (Bretz Guby, 1994, Stadelmann, 2000). Some adaptive traits of plant species are correlated with the conditions encountered in extremely disturbed situation such as rock slides. They demonstrate the strong role of processes occurring on the soil's surface (Körner, 2003; Käsermann, 2003). Moreover, particularly in alpine landscape, snow cover patterns can also exert pervasive ecological effects. Hence, the spatial and temporal persistence of the snow cover is strongly influenced by topography and determines species composition and spatial vegetation patterns (Walsh *et al.*, 1994). Snow cover affects also soil and vegetation moisture levels and offers protection from climatic stress, particularly wind desiccation (Schaefer, 1995). But snow could also be a stress factor itself if considering its dynamic aspect like avalanche paths or wind-induced snow translocation (Körner, 2003).

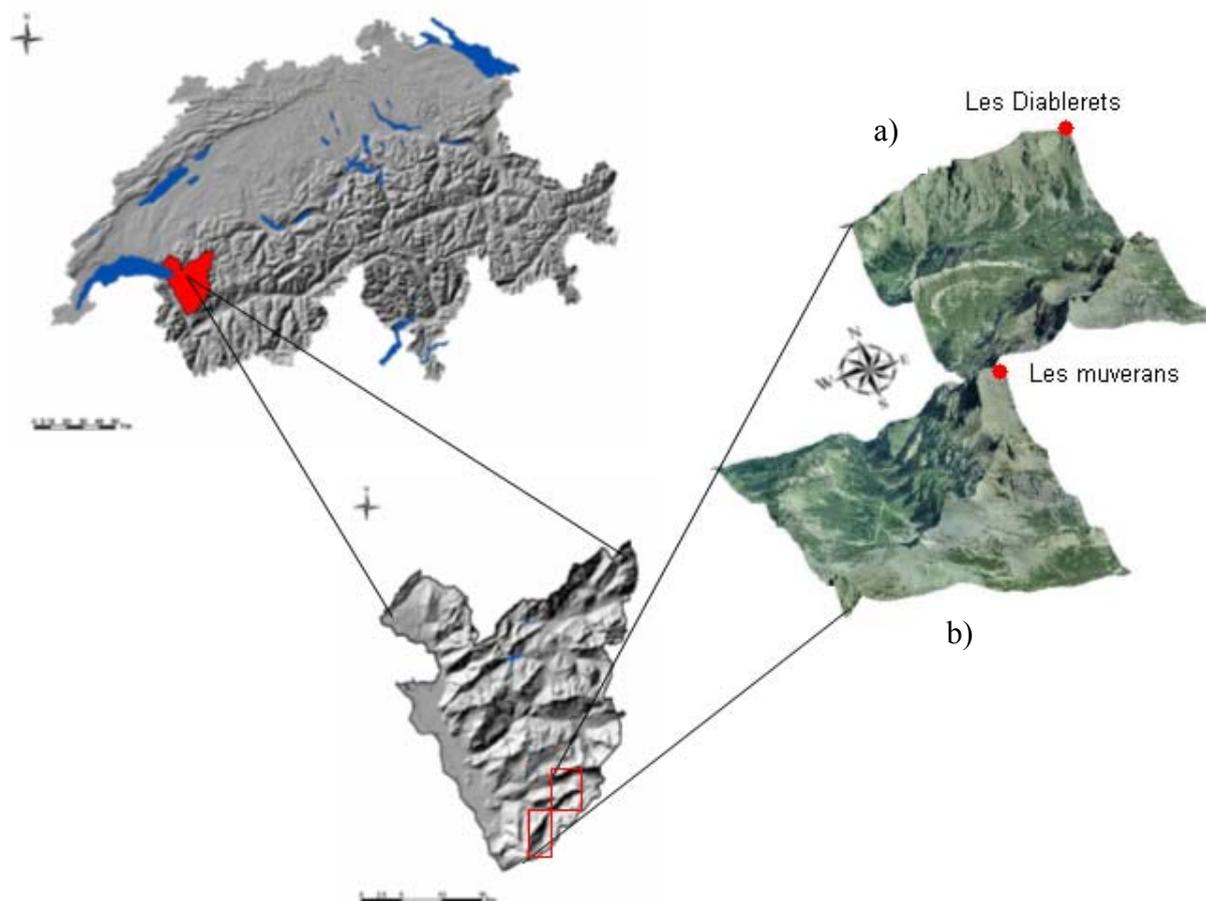
In this study we will integrate process-based disturbance variables in models of plant potential distribution (figure 1). In order to obtain such variables, we will transform the vegetation relevant information contained in geomorphological maps into a perturbation index. We will finally integrate process-based disturbance predictors like perturbation index and snow accumulation pattern into the static models to see how they could improve the predictive power for sub-alpine and alpine plant species.



**Figure 1:** Plant potential distribution is influenced by a wide range of variables at different levels. The perturbation intensity resulting from geomorphologic processes (like solifluxion or rock fall) is a variable that influence plant growth, development and potential distribution at the regulator level. (Adapted from Guisan & Thuiller in review)

## Material & methods

The study area is located in the swiss prealps. The area is named “Anzeindaz-Solalex” and forms a large plate at a mean elevation of 1900m. This area is surrounded by two big mountain chains (les Diablerets 3210m and les Muverans 3051m) and represents typically a calcareous alpine landscape. The vegetation is distributed according the typical climax of sub-alpine and alpine belts vegetation. With a surface of approximately 20km<sup>2</sup>, the altitudinal gradient extends on more than 1000 meters with various exposed faces (figure 2 a).



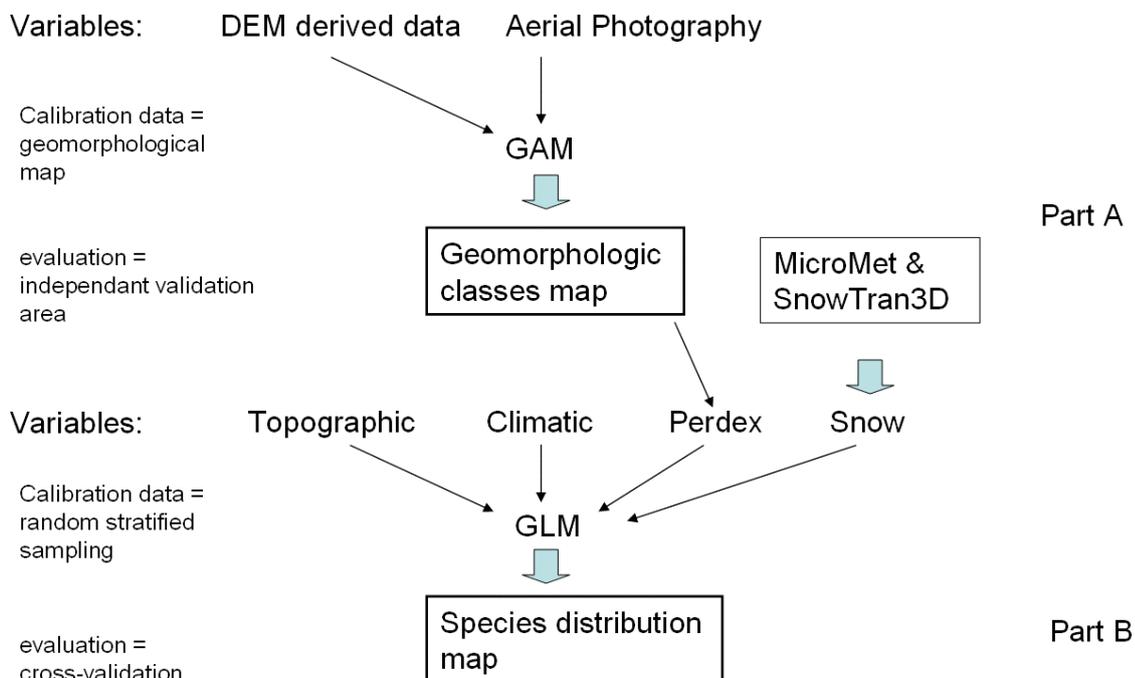
**Figure 2:** Geographic location of the study area, in red the MODIPLANT study area  
a) Anzeindaz-Solalex  
b) Vallon de Nant

Poorly affected by anthropic activities, the study area reveals a great bio-diversity with 382 plant species, which counts for the sixth of the swiss flora. This area is also well known from geographers because of to the wide range of geomorphologic processes occurring. The

geographic institute of the University of Lausanne (IGUL<sup>1</sup>) studies this area from ages and provides very detailed maps of the area (Rappaz & Hunziker, 1996; Loyer & Pahud, 1996). A small valley in the vicinity, the vallon of Nant is used as independent evaluation area for testing the perturbation index model (figure 2 b). This area is similar to the calibration area of Anzeindaz as it represents well the alpine environment. However the vallon de Nant has geographic particularities and presents less aspect variations than the calibration area due to the unique south-north oriented valley. This area was chosen because of the availability of geomorphological maps (Phillips 1993).

*Modeling framework:*

Geomorphologic classes are transformed into a perturbation index (perdex) for being added as continuous predictive variable into the plant distribution models. The part A consists of building a statistic model with digital elevation model (DEM) derived data and classified aerial photographs to produce a perturbation intensity map. The part B concerns the plant species distribution modeling, where models including several predictive variables are compared together (figure 3). The models are then implemented into a GIS using ARC/INFO 8.3 software (ESRI inc. 1999-2002) to produce distribution maps.



**Figure 3:** Modeling framework: in the part A geomorphologic variables are obtained and prepared. In the part B species distribution maps are produced with GLM models including several variables sets

## A) Data preparation/obtention

### *Species data*

A random stratified sampling was performed during summer 2002 and 2003 in order to collect presence and abundance information on plant species. Altitude, slope, aspect and soil permeability were the stratifiers. A second random sampling was performed during summer 2004 focusing especially on geomorphologic units which were extra stratifiers. Presence/absence data were obtained in 8 by 8 m plots according the MODIPLANT sampling protocol. All the sampling points are separated from at least 100m for avoiding spatial autocorrelation. The 2002-3 field campaign provided 41 plots and the summer 2004 field campaign provided 31 additional points. The final number of plots was 72 (table 1). 382 species were count during the sampling campaign but in this study we dropped rare species (*i.e.* species with less than 30 presence points) avoiding loss of robustness (Lehman *et al.*, 2002) and conserved only 92 species. The independent evaluation area does not provide a sufficient number of sampling plots (28) for assuming statistical robustness so it can not be used as independent validation area for the species model predictions.

**Table 1.** Distribution of the sampling points trough geomorphologic units and time.

	<b>2002-2003 samples</b>	<b>2004 samples</b>	<b>Total samples</b>
Glacial accumulation	16	0	16
Gravity accumulation	7	0	7
Lapiaz	5	1	6
Dejection cone	2	2	4
Rock fall	5	0	5
Avalanche path	0	6	6
Rock slide	3	16	19
Gelifluxion	0	3	3
Others	3	3	6
Together	41	31	72

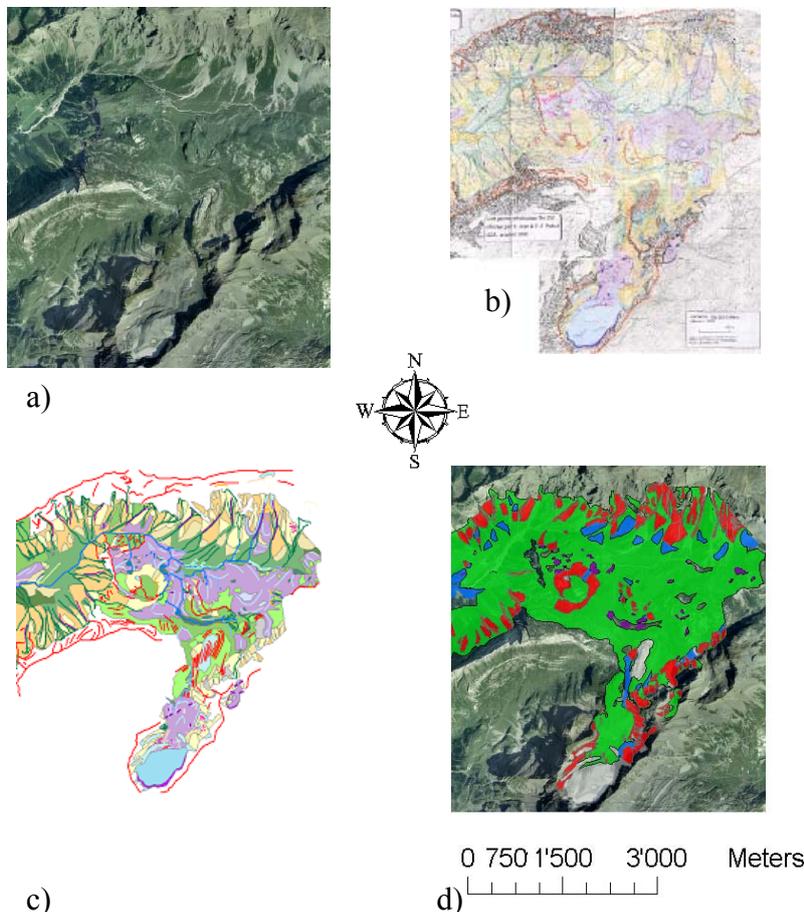
### Perturbation index

In a preliminary study (Spahr & Vuissoz 2004), we showed with co-variance analyses that the geomorphologic units as described by geographers could explain variance that was not correlated with environmental predictors usually used by ecologists. Those results also suggested that the geomorphologic units could be merged in three classes characterized by increasing perturbation intensity (table 3). The geomorphologic qualitative classes were transformed in a quantitative predictor which we named « perdex » for perturbation index with the following procedure:

The geomorphologic units are merged into three classes according their impact in terms of perturbation and digitalized into a GIS (figure 4).

**Table 3:** geomorphologic units as classified by covariance analyses

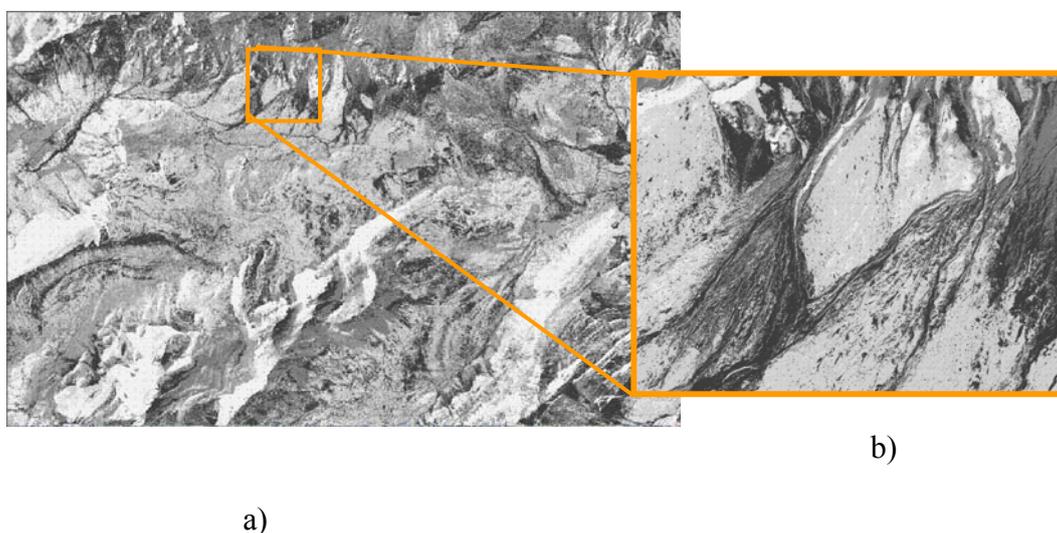
<b>High perturbation</b>	<b>Medium perturbation</b>	<b>Low perturbation</b>
Avalanche paths Rock slide Gelifluxion	Rock fall	Gravity accumulation Glacial accumulation Dejection cone Lapiaz



**Figure 4:**

- a) aerial view of the calibration area.
- b) geomorphologic map of the area.
- c) digitalized geomorphologic map
- d) geomorphologic units merged in three classes (*red*: high perturbation, *blue*: medium perturbation, *green*: low perturbation)

A Generalized Additive Model (GAM) is calibrated with topographic and aerial photographs in order to fit at best each of the three merged geomorphologic classes of the study area. Topographic predictive variables are derived from the digital elevation model 25m of the study area with an aml function (Reuter, 2003). Aerial orthophotographs of the study area with 0.5m by 0.5m pixel resolution are provided by Swisstopo<sup>1</sup>. Aerial othophotographs are then merged together in a mosaic picture and a basic image enhancement is performed with contrast and luminosity modifications, finally to improve the homogeneity a 3 by 3 pixel moving window low pass was applied. The unsupervised classification in five categories is executed with IMAGINE 8.6 (ERDAS 1991-2002) software (figure 5). Finally the resolution is downscaled to 25 meter pixel to fit the resolution of the DEM derived variables.



**Figure 5:** a) ortho photography of the area enhanced and classified in 5 categories (gray scale).  
b) detailed view of an area focusing on a rock slide surrounded by gullies

The software used to perform the GAM is SPLUS 6.2 (2000, MathSoft, INC.) with the GRASP 2.5 (Generalized Regression Analysis and Spatial Prediction, Landcare research 1999-2004) package. The input formula is of the type:

Geomorphologic classes  $\sim$  Logit(  $\alpha(j)$ Climatics +  $\alpha(j)$ topographics +  $\alpha(j)$ aerial classification)

<sup>1</sup> <http://www.swisstopo.ch/>

This GAM procedure allows four smoothing degrees of freedom for the additive function and the variable selection is performed by a both-sided stepwise selection with AIC. As the response variable is of presence/absence type, a binomial family distribution rule with logit link is used. Three models are built, one for each perturbation intensity (high perturbation, medium perturbation and low perturbation). The predictive variables involved in the stepwise selection are explicated in the table 4. Each model is calibrated within the study area and is also predicted in the independent evaluation area the vallon de Nant.

**Table 4:** Name, description and unit of the DEM derived variables retained to build the perturbation intensity models

<b>Name</b>	<b>Description</b>	<b>unit</b>
op5c	ortho photo enhanced and classified in five classes	unitless
Dem25	digital elevation model	[m]
plan curv	plan curvature	[1/100m]
profile curv	profile curvature	[1/100m]
fdir	direction of flow	ordinal [0-64]
flacc	accumulation flow to each cell	cells
ptcg1	position in landscape based on bassin	%
slp	slope	°

The high, medium and low models are then merged into a quantitative “perturbation index” predictor. The following weights are attributed to each perturbation model: +3\*high; +1\*medium; -1\*low.

It is then standardized to have values between zero (low perturbation) and one (high perturbation). The performance of the perdex layer is checked by counting the number of pixels correctly predicted into the calibration and evaluation area after having the probability of occurrence reclassified into presence/absence with the kappa max threshold (see below).

*Snow accumulation index (MicroMet and SnowTran-3D):*

A snow accumulation index layer is obtained with the courtesy of Glen Liston<sup>2</sup>, who provided us snow accumulation model based on the DEM of the area and an average dominant wind. Local wind direction and intensity is calculated in a quasi-physically-based meteorological model (MicroMet; Liston in press). Then a 1m uniform snow layer is applied to the area and a physically-based numerical snow transport model (SnowTran-3D; Liston and Sturm 1998) is applied to see where snow will accumulate and where it will disappear. Snow was blown

---

<sup>2</sup> Glen E. Liston, Ph D. Department of Atmospheric Science. Colorado State University

during 5 days with an average dominant wind of 280° at 15m/s. The resulting map is a snow accumulation index that provides information about the places where snow accumulates and places where snow disappears unlike the classical snow cover maps.

### *Climatic and topographic variables*

Environmental variables used to predict species distribution are hereafter called predictors (table 2). The resolution of the environmental maps is 25m\*25m per pixel. Topographic data come from the Federal Office of Topography (Swisstopo); climatic data come from the Swiss Federal Institute for Forest, Snow and landscape Research (WSL<sup>3</sup>).

**Table 2:** name, description and unit of the climatic and topographic variables used to build species distribution models

<b>Name</b>	<b>Description</b>	<b>unit</b>
Srad7	sum of july global potential short wave radiation	[KJ/day]
Ddeg0	Annual degree days above 0°C threshold	[Degree*Day]
Topo	topographic position between 50m and 1550m with 250m increments moving windows size	unitless

## **B) Building the models**

The species distribution models are calculated using an automated loop SPLUS which calibrates a Generalized linear model (GLM; McCullagh and Nelder, 1989). GLM's are fitted with a binomial distribution and a logistic link function and consider single and second term polynomials. The selection of the climatic variables is performed with a two-sided stepwise procedure with AIC.

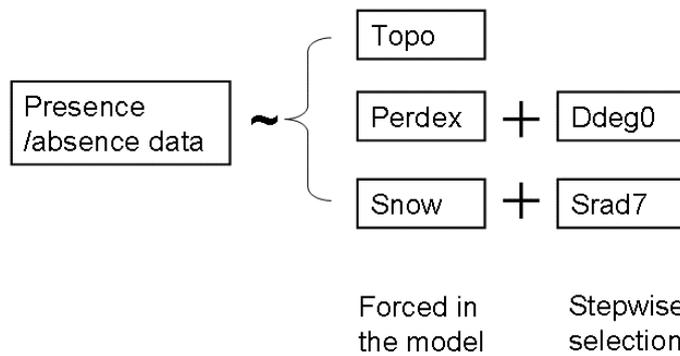
For each species four models are built.

First the null model contains as input formula:

Presence / absence data ~ logit link function( $\alpha$  Ddeg0 +  $\beta$  Srad7)

Then the snow model includes the snow accumulation index as predictive variable, the perdex model adds the perturbation index and the topo model includes the topographic position index. Snow, perdex and topo variables are forced into the models bypassing the stepwise selection procedure (figure 6).

<sup>3</sup> <http://www.wsl.ch/welcome-en.ehtml>



**Figure 6:** for each species four models are built, three with topo-climatic and process-based disturbance variables and one only with topo-climatic variables

A potential distribution map is then obtained from the estimation of the coefficients of each retained predictor and scaled with a logit link function transformation in order to fit the 0-1 scale of the presence/absence answer variable.

### C) Evaluating the models

#### *Adjusted explained deviance (adj. $D^2$ )*

This is the percentage of variance explained by the GLM. This measure expresses the fit of the model (Guisan, Weiss & Weiss 1999). Adjusted explained deviance has a range of variation from 0 to 1 where 1 means the perfect fit of the calibration data. A big value indicates that the model fit well the actual distribution of the species. The adjusted deviance is the explained deviance corrected in regard of the amount of degrees of freedom. As we are specifically interested in the amount of variation in the presence/absence pattern of species depending on the set of predictors used, the adj.  $D^2$  is a good measure.

### *Area under the curve value (AUC)*

As we have enough presence/absence data available only for a small alpine area in the MODIPLANT study zone, a quasi-independent data set is used for evaluation. A stratified cross validation statistic with ten groups was performed, nine groups serve as calibration data and the tenth is evaluated (Ecospat library, Guisan & Randin, 2004). The threshold-independent Receiver Operating Characteristic (ROC) approach was obtained by calculating the area under the roc curve. AUC value is a measure of the accuracy and quality of the model as it calculates the prediction success. Bigger is the area under the ROC-plot curve best the model suits. Even if there is no best threshold for this measure it is useful to compare models together (Fielding, A. H. and Bell, J. F. 1997). The Gini coefficient is calculated:  $AUC' = 2 * (AUC - 0.5)$  in order to have values ranging from zero (for an uninformative model) to one (for a model with perfect discrimination) (Hand & Henley 1997).

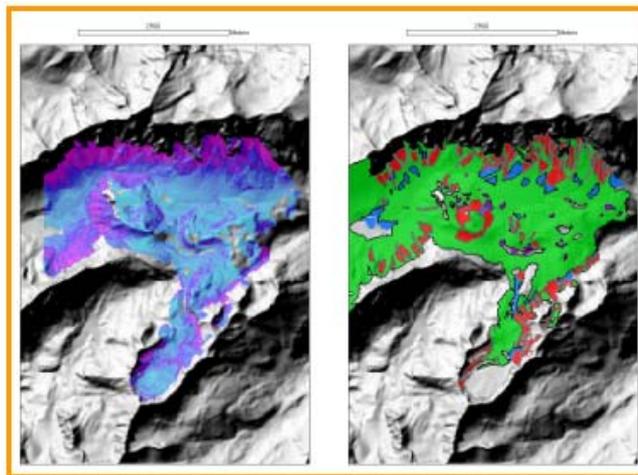
### *Best kappa (B-kappa)*

The kappa coefficient (Cohen 1960; Fielding & Bell 1997) was calculated for all thresholds between zero and one by increments of 0.05. The greatest value was kept as the 'best kappa' value (Elith 2002). This measure expresses the best possible agreement not obtained randomly between two qualitative variables (of which a binary variable is a particular case). Then the best kappa threshold is estimated to constrain adjusted probabilities of occurrence into binary presence absence on the predictive distribution map.

## Results

### *Perturbation index:*

The GAM procedure combining aerial photographs and topographic variables provided surprisingly accurate maps at first sight (figure 6). However adjusted explained deviance and AUC' of the high, medium and low perturbation models are very weak (table 5). High perturbation areas, represented in purple on the perdex map, match up in most cases the high perturbation areas described by geographers. Low perturbation areas, predicted in light blue, occupies nearly the same spaces as the green, low perturbation areas described by geographers.

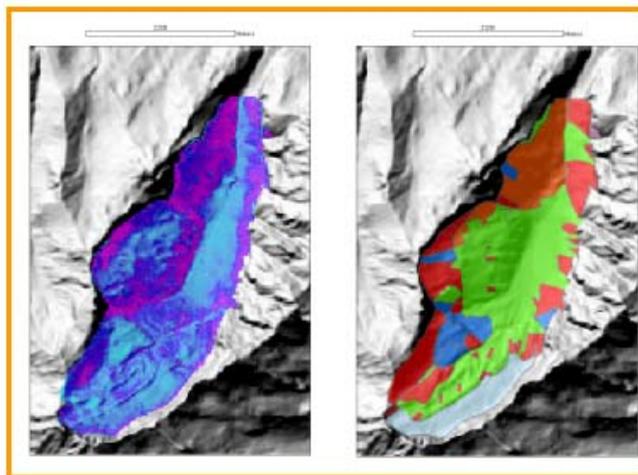


a)

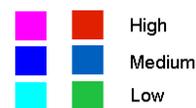
### **Figure 6:**

a) calibration area: visual evaluation of the perdex model map (left) and classified geomorphologic map (right).

b) evaluation area : visual match between the perdex model map (left) and the geomorphologic map (right)



b)



**Table 5** : adj. D<sup>2</sup>, AUC, AUC' and max kappa threshold of the perturbation models

	adj. D <sup>2</sup>	AUC	AUC'	Max kappa threshold
high perturbation	0.31	0.727	0.454	0.25
medium perturbation	0.28	0.692	0.416	0.20
low perturbation	0.26	0.771	0.542	0.35

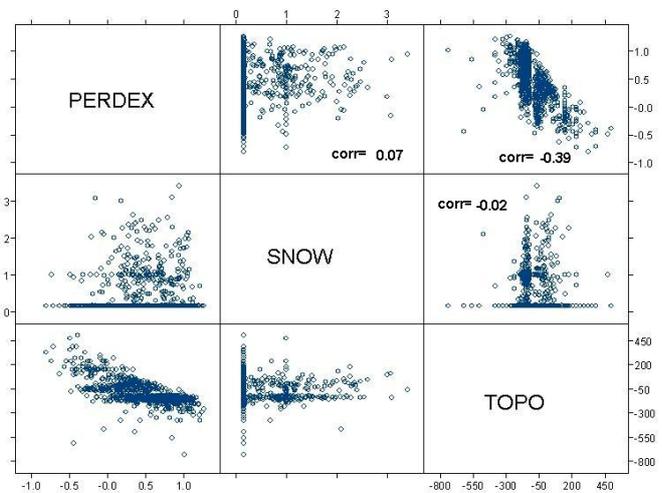
Prediction of the perturbation intensity into other areas was evaluated with another available geomorphological map for the vallon de Nant area. The proportion of well predicted pixels is quite similar between the calibration and the evaluation area (54.9% and 48.4 %) and there is a trend to overestimate the intensity of the perturbation (table 6). The counting of well predicted pixels within the calibration and evaluation area (table 6) shows that the models tends to overestimate the perturbation intensity and that the good predictions are rare even if better than random.

**Table 6** : evaluation of the perturbation index model

	Calibration area	Evaluation area
total number of pixels	17462	11022
good predicted pixels	9579	5340
% good predictions	0.549	0.484
% underestimated	0.149	0.230
% overestimated	0.303	0.286

The comparison of the perdex layer to other variables used in the models shows that there is no correlation between them (figure 7) so they could not at first sight explain the same part of deviance in the models.

Snow and perdex variables have a correlation index of 0.07, snow and topo share a -0.02 correlation index and the most correlated of the predictors here are perdex and topo which have a -0.39 correlation. However the distribution of the values looks homogeneous with few outliers values.



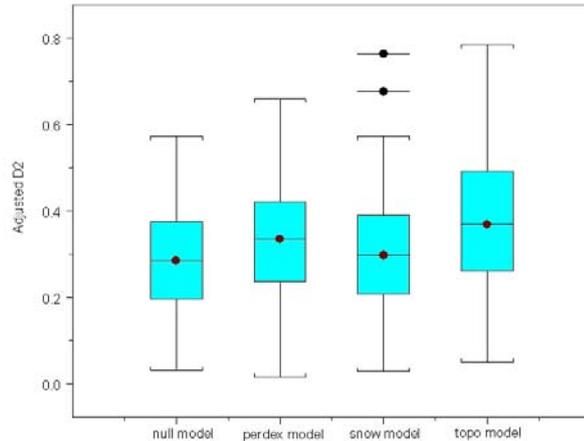
**Figure 7**: correlation matrix of the variables perdex snow and topo.

*Species distribution modeling:*

GLM were fitted for 92 species with four types of models. Globally all models show significant differences with each other if we compare their adj.  $D^2$  with paired t.test with the other variables (table7 and figure 8). The perdex and snow model improve slightly the average of adjusted  $D^2$  and the average of AUC' but the topo model is still the best except for the AUC'.

**Table 7:** comparison between the four different models

Model name	Mean adj. $D^2$	Mean AUC'	p-value <0.05
null	0.29	0.52	Yes
snow	0.31	0.51	Yes
perdex	0.33	0.58	Yes
topo	0.38	0.56	Yes



**Figure 8:** box-plot representation of the adj.  $D^2$  of the different models

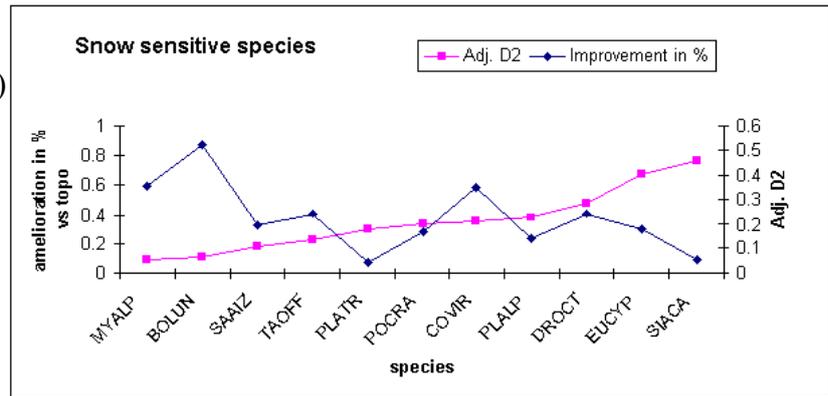
Models including topo variable globally suit the best various species distribution. But it appears that many species are more sensitive to “snow” or “perdex” variable than to topo predictor. It seems that several models are better to predict species occurrence with snow or perdex variable than with topo predictor. In this case we compared the adjusted  $D^2$  between snow/perdex and topo models in aim to see how many times snow/perdex models showed an improvement greater than 5% of the deviance explained by the topo model (table 8). Eleven snow models improved the prediction at the above given threshold (figure 9a). Sixteen perdex models showed such improvements (figure 9b). See annex 1 for species full names.

**Figure 9:**

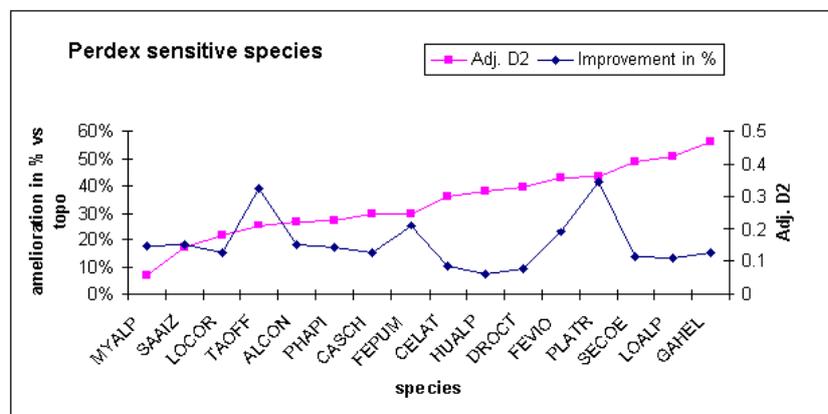
a) snow sensitive species are species that show an improvement of more than 5% compared to the adj.  $D^2$  that they had with topo variable instead of snow.

b) perdex sensitive species are species whom adj.  $D^2$  was raised by more than 5% compared to the adj.  $D^2$  they had with topo variable instead of perdex

a)



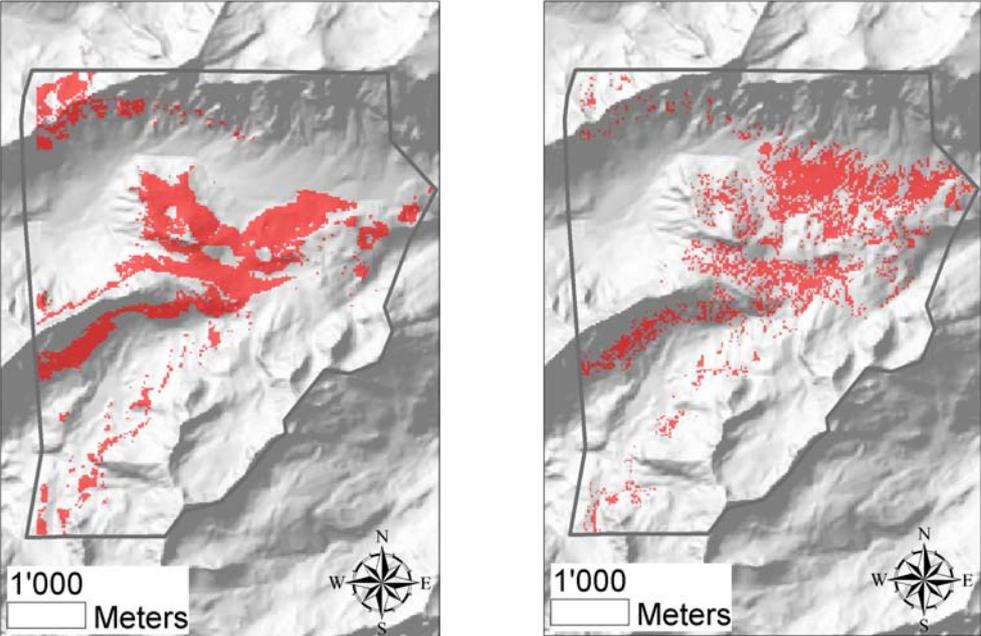
b)



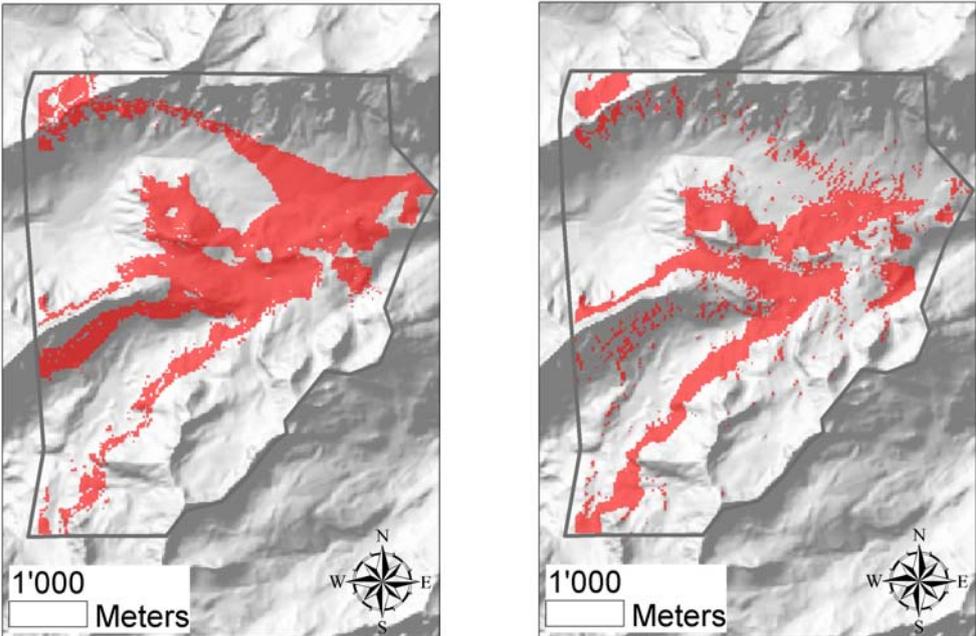
**Table 8:** perdex sensitive and snow sensitive species with adj.  $D^2$ , Auc', threshold kappa and increase of the adj.  $D^2$  in % in regard of topo model

Perdex sensitive:	Adj. $D^2$	Auc'	Treshold kappa	increase	Snow sensitive	Adj. $D^2$	Auc'	Treshold kappa	increase
ALCON	0.26	0.49	0.35	15.04%	BOLUN	0.10	0.13	0.2	52.34%
CASCH	0.29	0.55	0.55	12.75%	COVIR	0.35	0.52	0.35	35.23%
CELAT	0.36	0.55	0.55	8.59%	DROCT	0.47	0.65	0.40	24.06%
DROCT	0.39	0.62	0.60	7.87%	EUCYP	0.67	0.79	0.50	17.87%
FEPUM	0.3	0.53	0.25	21.33%	MYALP	0.09	0.13	0.30	35.87%
FEVIO	0.42	0.55	0.40	19.11%	PLALP	0.37	0.71	0.35	14.17%
GAHEL	0.56	0.82	0.30	12.43%	POCRA	0.33	0.66	0.40	17.01%
HUALP	0.38	0.70	0.30	6.04%	SAAIZ	0.18	0.21	0.40	19.34%
LOALP	0.50	0.71	0.50	11.09%	SIACA	0.76	0.51	0.20	5.37%
LOCOR	0.21	0.51	0.25	12.50%	TAOFF	0.23	0.47	0.20	24.56%
MYALP	0.06	0.24	0.25	14.49%					
PHAPI	0.27	0.57	0.45	14.39%					
PLATR	0.43	0.70	0.45	34.71%					
SAAIZ	0.17	0.37	0.25	15.12%					
SECOE	0.48	0.68	0.50	11.25%					
TAOFF	0.25	0.43	0.35	32.55%					

The homogenous distribution patterns provided by the topo models are more fragmented in the snow and perdex models. For instance, *Coeloglossum viride* has an adjusted deviance rising from 0.23 to 0.35 with the help of the snow accumulation variable. It also has its potential distribution restricted by high amount of snow accumulation (figure 10). The distribution map of *Dryas octopetala* (figure 11) shows on the contrary new occupied small areas above the upper topo-climatic limit in the vicinity of avalanche paths.

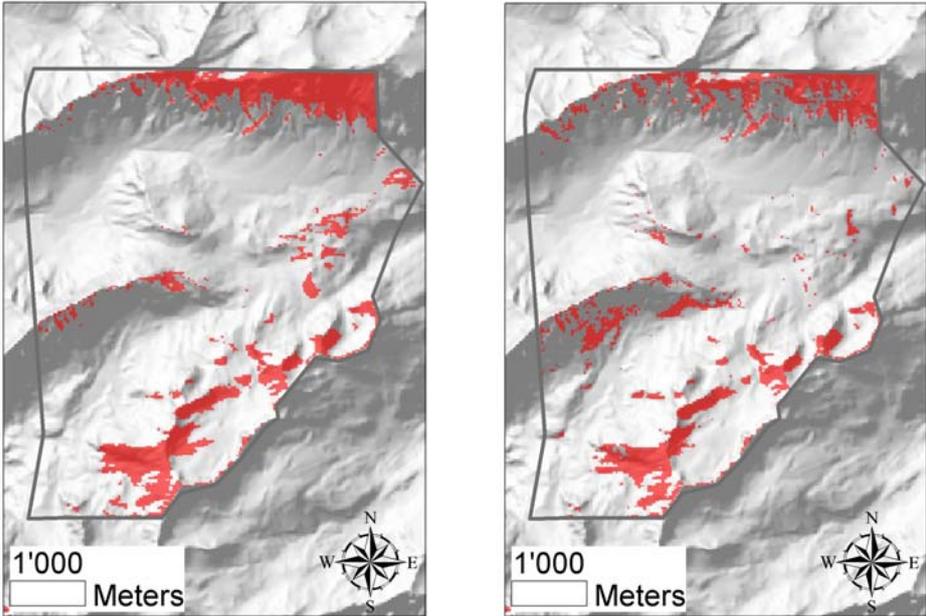


**Figure10 :**  
*Coeloglossum viride* distribution map. The topo model on the left has a 0.23 adj. D2 and the snow model on the right gives 0.35 adj. D2

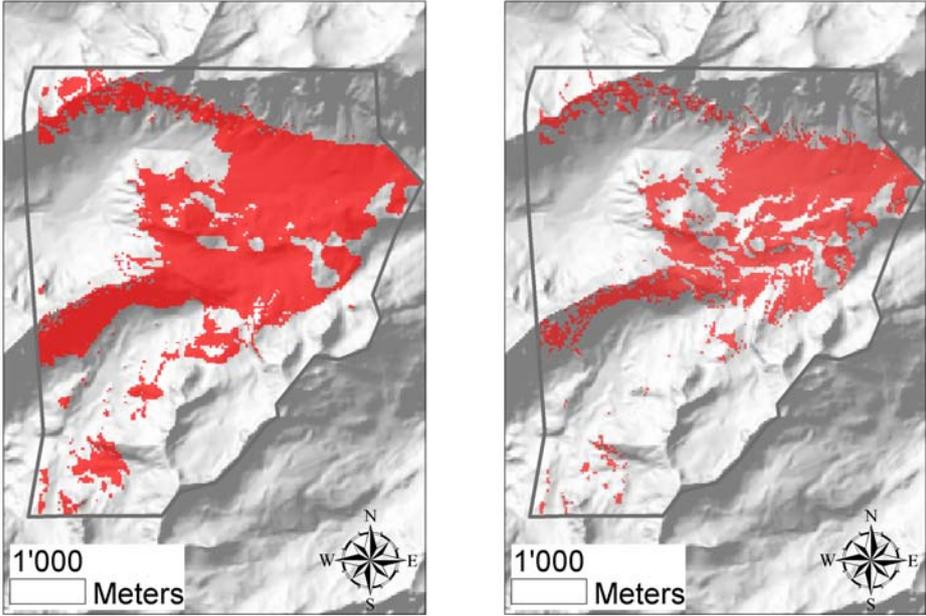


**Figure11 :**  
*Dryas octopetala* distribution map. The topo model on the left has a 0.23 adj. D2 and the snow model on the right gives 0.35 adj. D2

Potential distributions maps produced with perdex model shows similar area fragmentation than snow model. The example of *Festuca violacea* distribution map shows that the perdex model modifies the distribution patterns especially on big south faced high altitude slopes (figure 12). *Plantago atratum* shows similar pattern changes with an alteration in the contrast between high and low probabilities occurrence.



**Figure 12:**  
*Festuca violacea* distribution map. Topo model on the left provides 0.34 adj. D2 and the perdex model on the right provides 0.43 adj. D2



**Figure 13 :**  
*Plantago atrata* ditribution map. Topo model on the left provides 0.28 adj. D2 and perdex model on the right gives 0.43 adj. D2

## Discussion

Building plant species distribution models requires particular attention especially when performing the choice of the predictors. The quest of new predictors seems to be one of the ways to improve plant distribution models. The results we presented show that process-based disturbance variables can explain a significant part of deviance that is not explained by topoclimatic variables classically used. First we will discuss about the perturbation index as a new predictive layer, and then we will analyze the general impact of snow accumulation and perturbation index variables. Finally the specific case of *Dryas octopetala*, and *Festuca violacea aggr*, will be used to illustrate two particular examples where snow and perturbation index improve dramatically the small scale predictions.

### *Perturbation index:*

The perturbation index built in this study with a combination of visible range of aerial photographs and topographic data within a GAM procedure seems to be an interesting way to deal with geomorphologic processes. The use of visible range aerial photographs is justified by the fact that geographers itself rely in part on visual information to build the maps. Data required to set up such a layer are easy to obtain and are available at various scales and time period. Aerial photographs are for instance often brought up to date and the resolution grows regularly as costs are going down.

However if we look at the explained deviance or validation measurements like AUC', the perturbation models appears to have a low accuracy to the geomorphological map. It could mean that the perdex map is a bad indicator of perturbation intensity. But we don't think so because the produced maps look really good and suit well which was observed on the field. The bad evaluation scores of theses models may come from the way which geomorphological maps are drew. Geomorphologic processes are delimited on the map by large polygons and there is no transition zone between two distinct neighboring geomorphologic units. Therefore it is very hard to fit it with topographic variables in a GAM.

The intensity of geomorphology phenomena is not homogenous in space and in time. However the abrupt limits drawn on geomorphological maps do not exists in reality. Moreover, the geomorphologic processes influence species distribution continuously

according to their intensity. Considering this aspect of the geomorphologic impact, the use of a continuous perturbation index instead of a semi-qualitative geomorphologic predictor roughly extracted from the geomorphological map enables to build more direct predictors as the plant response to perturbations is not semi qualitative. This fact let us think that the perturbation index is more appropriate for predicting the distribution of species than the crude geomorphologic units.

More advanced remote sensing methods like contextual merging (Romstad 2002) or pattern recognition (Barlow 2002) may be used to determine more precisely the intensity of perturbation. For instance the size and the spatial organization of the stones along a rock slide gives information about the intensity and the frequency of the perturbation events. However an expert approach based on advanced remote sensing knowledge could have lead to better results (Brown 1998) using multi-spectral bands information or fuzzy logic classification. Beyond, time imperatives and gaps in remote sensing knowledge forced us to go ahead in this study preventing systematic testing and procedure optimization.

#### *Global assessment, topo vs snow/perdex:*

On the whole, if we look at the explained deviance of the models, the differences between the models are low and they do not justify moving one particular variable away instead of another except for the topo variable that provides very good results with many species. However some species like *Coeloglossum viride*, *Dryas octopetala*, *Festuca violacea aggr.*, *Plantago atrata s.str.* show a particular sensitiveness to the snow and perdex variables. Using theses process-based disturbance variables improve significantly the models for theses species even if the general shape of the distribution maps does not vary greatly according the choice of the predictor. A closer look at the distribution maps reveals that differences are more relevant at the micro-scale. The large and homogenous areas predicted with topo-climatic variables appear to be splitted into smaller fragmented favorable areas when using snow or perdex predictors. The hydro mechanicals predictors tested in this study explain certainly a valuable part of additional explained deviance but large and meso-scale models may take more benefit from variables influencing larger areas like land use occupation (Dirnböck 2000; Dullinger 2004; see Jacquard 04).

*Two examples of punctual model amelioration:*

Model improvement due to snow variable is particularly astonishing if we consider the case of *Dryas octopetala*. The place of *Dryas octopetala* within the vegetation units is between pioneer and dry meadow species and it is able to colonize places just after early pioneer species (Dutoit 1981, Randin 2002). This species, also known as alpine dryad, resists to extreme gradients of temperature (Körner 2003) and strong winds. For these reasons it is often found on rocks crops with poor organic accumulation. This species responds particularly well to the snow variable because these rocks that it colonizes are places where snow accumulation is rapidly displaced due to wind and radiation exposure. Important solar radiation income and very high temperature episodes occurring on such places are typically ecological conditions which are not described by the climatic predictors usually used like incoming radiation or temperature average measurements.

*Festuca violacea* is a Poaceae which forms a compact tuft of fine metallic turquoise-blue leaves. It responds particularly well to the perturbation index variable. Disrupting soils conditions encountered on gully erosion can for example prevent competing species to establish. Here the perturbation index layer operates like a filter for unsuitable habitats. Potential distribution of *Festuca violacea* is more reduced at high altitude by the frequent perturbation along big slopes like rock falls than by the slope itself. *Festuca violacea* has been found in the field in the surroundings of high disturbed areas where only species from the *Thlaspion* and some other lithophile species can persist. Neither topographic position nor slope variable can discriminate *Festuca violacea* or *Thlaspi repens* habitat but it seems that a perturbation index can do it.

*Small scale predictive improvement and consequences:*

Small scale distribution observation can lead to a more comprehensive realized niche of species, where local influence of dynamic processes may have strong repercussions on distribution modeling. When the statistical models are applied to predicting species distribution under changing climate scenarios they show certain sensitivity to the initial conditions. Hence, considering the fine processes leading to modifications of the species niche, we can preserve modeling from systematic errors to spread through modeling cycles. In

the case of migration modules combined to statistical models like MIGRATOR (Engler, in press), the characterization of unfavorable areas as dispersion barriers could boost the accuracy of the predictions.

*Perspectives:*

Generalization of the process-based disturbance predictors to other alpine areas is possible as we showed with the vallon de Nant area. However, systematic testing about the best way to set up a perturbation intensity variable is needed and the use of remote sensing methods like spatial or spectral enhancement is surely able to bypass limitations encountered with aerial photographs.

Besides geomorphological maps describe dynamic processes that have a big impact on plant species distribution and it is a valuable source of information that botanists have at disposal.

Statistical models were fitted using species occurrence (presence absence). As this latter was derived from abundance data, in further studies models should be fitted using directly density instead of presence absence. This could sensibly change the results because the abundance is more relevant to the equilibrium hypothesis than simple presence/absence data that could be due to a temporary perturbation. However in the disturbed areas that we sampled, species density was quite low so we don't expect very different results.

### *Acknowledgments:*

First I want to thank Antoine Guisan for proposing this diploma subject. I am grateful to all the MODIPLANT group, Phd's, students for the great time I spend in the SIG lab and especially to Christian Parisod, Stéphanie Spahr and Christophe Randin who guided me through rock slides, fresh snow and spatial ecology. Finally, many thanks to Glen Liston who ran the snow transport simulations.

### *References:*

- Bakkenes, M., Alkemade, J.R.M., Ihle, F., Leemans, R. and Latour, J.B., (2002). Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology* 8, 390-407.
- Barlow J.,(2002). Satellite and Digital Terrain Data in the Detection of Landslides American Geophysical Union, Fall Meeting 2002
- Bretz-Guby, N., (1994). Géomorphologie et végétation à l'étage alpin: L'exemple du Mont-Gelé. Mémoire de l'Institut de Géographie, Faculté des lettres, Université de Lausanne.
- Brown D.G., (1998). Supervised classification of types of glaciated landscapes using digital elevation data. *Geomorphology* 21 233-250
- Cohen J.,(1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Measurements* 20, 37-46.
- Dirnböck T., Grabherr G., (2000). GIS Assessment of Vegetation and hydrological change in high mountain catchment of the northern limestone alps. *Mountain research and development* vol 20 No2: 172-179
- Dullinger S., Dirnböck T., Grabherr G., (2004). Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. *Journal of Ecology* 92, 241-252
- Elith, J. & Burgman, M., (2002). Predictions and their validation: rare plants in the central highlands, Victoria, Australia. *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds J.M. Scott, P.J. Heglund, J.B. Hafler, M. Morrison, M.G. Raphael, W.B. Wall & F. Samson), pp. 303–313. Island Press, Covelo, CA.

- Fielding, A. H. and Bell, J. F., (1997). A review of methods for the assessment of prediction errors in conservation presence / absence models. *Environmental Conservation* 24: 38-49
- Guisan, A., Theurillat, J.-P. and Kienast, F., (1998). Predicting the potential distribution of plant species in an alpine environment, *Journal of Vegetation Science* 9: 65-74.
- Guisan, A. and Theurillat, J.-P., (2000). Equilibrium modelling of alpine plant distribution and climate change : how far can we go ? *Phytocoenologia* 30: 353-384.
- Guisan, A., Weiss, S.B. & Weiss, A.D., (1999). GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, 143, 107–122.
- Hand, D.J. & Henley, W.E., (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society A*, 160, 523–541.
- Hirzel, A. & Guisan, A., (2002). Which is the optimal sampling strategy for habitat suitability modelling ? *Ecol. Modelling* 157(2-3): 331-341
- Käsermann C. Meyer F. Steiner A., (2003). Le monde végétal de Zermatt. Département des transports, de l'équipement et de l'environnement du canton du Valais. Rotten Verlags AG Visp 245 pp.
- Körner C., (2003). Alpine plant life: functional plant ecology of high mountain ecosystems. 2<sup>nd</sup> edition Springer 332 pp.
- Lehmann, A., Overton, J.McC. & Leathwick, J., (2002). GRASP:generalized regression analysis and spatial prediction. *Ecological Modelling*, 157, 189–208.
- Loye, E., Pahud, P.-F., (1996). Levé géomorphologique de Solalex-Anzeindaz. Travail de certificat de l'institut de géographie No250, Facultés des Lettres, Université de Lausanne.
- McCullagh, P. & Nelder, J.A., (1989). *Generalized Linear Models*, 2nd edn. Chapman & Hall, London, UK.
- Nichols W. F. Killingbeck K. T. August P.V., (1998). The influence of Geomorphological heterogeneity on biodiversity. *Conservation Biology* (12) 2
- Phillips Marcia, (1993). Géomorphologie du Vallon de Nant. Travail de Mémoire de l'Institut de géographie No331, Facultés des Lettres, Université de Lausanne.
- Randin C. ,(2002). Etude diachronique d'une zone alluviale dans le Vallon de Nant (Préalpes vaudoises) Travail de diplôme. Institut d'écologie Université de Lausanne.
- Rappaz, F.Hunziker, J.-l., (1996). Levé géomorphologique d'Anzeindaz. Travail de certificat de l'institut de géographie No245, Facultés des Lettres, Université de Lausanne.
- Reuter I. H. ,(2003). TOPO – ein Algorithmus zur reliefanalyse, IALE-region deutschland Jahrestagung 2003 in Eberswalde 71.
- Romstad B., (2001) Improving Relief Classification with Contextual

Merging Proceedings of the 8<sup>th</sup> Scandinavian research conference on geographical information science 2001

Schaefer, J. A., Messier F. , (1995). Scale-dependent correlations of arctic vegetation and snow cover. *Arctic and Alpine Research* 27: 38-43

Spahr S., Vuissoz G., (2004). Influence des variables géomorphologiques sur la végétation dans la région d'Anzeindaz. Travail de certificat de botanique. Université de Lausanne, non publié

Stadelmann C. ,(2000). Comparaison entre la géomorphologie et la végétation dans le Haut-Vallon de Réchy, IGUL, Mémoire de licence, non publié

Thuiller W., (2004). Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology* 10:12, 2020-2027

Walsh, S.J., Butler, D.R., Allen, T.R., Malanson, G.P., (1994). Influence of snow patterns and snow avalanches on the alpine treeline ecotone. *J. Vegetation Sci.* 5, 657–672.

Waser L., Catalini M., Schardt M., Schmitt U., Zini E., (2000). Inventory of alpine relevant parameters for an alpine monitoring system using remote sensing data. IAPRS, Vol XXXIII, Amsterdam, 2000

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

## Annex :

Annex 1: results of the GLM models:

Abrev.	Full name	Null		perdex		snow		topo	
		adj.D2	AUC'	adj.D2	AUC'	adj.D2	AUC'	adj.D2	AUC'
ADGLA	<i>Adenostyles glabra</i>	0.22	0.56	0.27	0.56	0.22	0.43	0.32	0.61
AGRUP	<i>Agrostis rupestris</i>	0.26	0.63	0.35	0.66	0.33	0.45	0.35	0.64
AGTEN	<i>Agrostis capillaris</i>	0.24	0.57	0.26	0.60	0.24	0.54	0.28	0.60
ALCON	<i>Alchemilla conjuncta</i>	0.18	0.44	0.27	0.49	0.16	0.27	0.23	0.47
ALCXA	<i>Alchemilla xanthochlora</i>	0.39	0.71	0.39	0.74	0.42	0.73	0.49	0.77
ALGLA	<i>Alchemilla glabra</i>	0.11	0.29	0.11	0.20	0.10	0.23	0.10	0.20
ANCHA	<i>Androsace chamaejasme</i>	0.32	0.58	0.34	0.50	0.33	0.52	0.41	0.63
ANTOD	<i>Anthoxanthum odoratum</i>	0.20	0.54	0.22	0.54	0.22	0.48	0.24	0.54
ANTVL	<i>Anthyllis vulneraria s.l.</i>	0.39	0.65	0.45	0.70	0.38	0.57	0.60	0.83
BAALP	<i>Bartsia alpina</i>	0.24	0.55	0.23	0.61	0.24	0.52	0.31	0.57
BEMIC	<i>Aster bellidiastrum</i>	0.26	0.56	0.26	0.48	0.25	0.51	0.34	0.54
BOLUN	<i>Botrychium lunaria</i>	0.03	0.14	0.02	0.02	0.11	0.13	0.05	-0.02
CABAR	<i>Campanula barbata</i>	0.20	0.48	0.30	0.41	0.18	0.35	0.52	0.57
CADEF	<i>Carduus defloratus s.str.</i>	0.19	0.49	0.20	0.46	0.17	0.36	0.25	0.45
CASCH	<i>Campanula scheuchzeri</i>	0.17	0.49	0.30	0.56	0.15	0.46	0.26	0.58
CASEM	<i>Carex sempervirens</i>	0.41	0.67	0.49	0.72	0.40	0.58	0.60	0.83
CASIM	<i>Carlina acaulis</i>	0.26	0.48	0.24	0.46	0.29	0.54	0.33	0.45
CELAT	<i>Cerastium latifolium</i>	0.32	0.61	0.36	0.56	0.31	0.53	0.33	0.56
CHPRA	<i>Leucanthemum vulgare</i>	0.25	0.67	0.30	0.65	0.26	0.57	0.29	0.52
CISPI	<i>Cirsium spinosissimum</i>	0.14	0.31	0.14	0.34	0.12	0.26	0.14	0.36
COVIR	<i>Crepis aurea</i>	0.21	0.38	0.24	0.40	0.35	0.52	0.23	0.22
CRAUR	<i>Deschampsia cespitosa</i>	0.49	0.80	0.51	0.79	0.51	0.73	0.53	0.82
DECAE	<i>Deschampsia cespitosa</i>	0.23	0.56	0.21	0.47	0.21	0.46	0.22	0.45
DROCT	<i>Dryas octopetala</i>	0.35	0.66	0.39	0.62	0.48	0.65	0.36	0.59
EUCYP	<i>Euphorbia cyparissias</i>	0.40	0.73	0.42	0.77	0.68	0.80	0.56	0.79
EUMIN	<i>Euphrasia minima</i>	0.26	0.60	0.25	0.47	0.26	0.56	0.43	0.64
FEPUM	<i>Festuca quadriflora</i>	0.20	0.28	0.30	0.53	0.18	0.26	0.24	0.47
FERUB	<i>Festuca rubra</i>	0.36	0.70	0.39	0.71	0.36	0.62	0.53	0.77
FEVIO	<i>Festuca violacea</i>	0.36	0.61	0.43	0.56	0.35	0.56	0.35	0.58
GAANI	<i>Galium anisophyllum</i>	0.35	0.64	0.35	0.69	0.34	0.60	0.52	0.70

GAHEL	<i>Galium megalospermum</i>	0.47	0.81	0.56	0.83	0.46	0.75	0.49	0.76
GEMON	<i>Geum montanum</i>	0.22	0.54	0.39	0.59	0.28	0.52	0.46	0.61
	<i>Gentiana campestris</i>								
GENCA	s.str.	0.18	0.52	0.25	0.47	0.20	0.46	0.36	0.51
GENPU	<i>Gentiana purpurea</i>	0.30	0.56	0.38	0.49	0.28	0.47	0.76	0.52
GEVER	<i>Gentiana verna</i>	0.36	0.68	0.41	0.67	0.44	0.68	0.49	0.76
GLCOR	<i>Globularia cordifolia</i>	0.53	0.80	0.54	0.82	0.54	0.78	0.52	0.72
GYREP	<i>Gypsophila repens</i>	0.29	0.60	0.27	0.48	0.27	0.53	0.38	0.62
	<i>Helianthemum</i>								
HELNU	<i>nummularium</i>	0.28	0.59	0.31	0.52	0.29	0.55	0.32	0.48
HEVER	<i>Helictotrichon versicolor</i>	0.42	0.68	0.45	0.64	0.41	0.61	0.64	0.79
HIBIF	<i>Hieracium bifidum</i>	0.13	0.30	0.10	0.30	0.15	0.33	0.15	0.41
HICOM	<i>Hippocrepis comosa</i>	0.34	0.57	0.38	0.51	0.33	0.41	0.50	0.52
HIEVI	<i>Hieracium villosum</i>	0.28	0.51	0.28	0.57	0.27	0.33	0.27	0.44
HIMUR	<i>Hieracium murorum</i>	0.35	0.63	0.36	0.64	0.35	0.55	0.37	0.61
HOALP	<i>Homogyne alpina</i>	0.32	0.64	0.36	0.65	0.30	0.58	0.55	0.78
HUALP	<i>Pritzelago alpina</i> s.str.	0.36	0.71	0.38	0.71	0.35	0.62	0.36	0.72
LEOHA	<i>Leontodon hispidus</i>	0.37	0.73	0.40	0.73	0.37	0.71	0.47	0.78
LIMUT	<i>Ligusticum mutellina</i>	0.16	0.51	0.17	0.36	0.15	0.43	0.22	0.48
LOALP	<i>Lotus alpinus</i>	0.41	0.73	0.51	0.71	0.39	0.65	0.45	0.71
LOCOR	<i>Lotus corniculatus</i>	0.17	0.37	0.22	0.51	0.15	0.36	0.19	0.43
LUMUL	<i>Luzula multiflora</i>	0.39	0.72	0.51	0.75	0.39	0.55	0.49	0.75
MYALP	<i>Myosotis alpestris</i>	0.09	0.20	0.07	0.24	0.09	0.13	0.06	0.08
NASTR	<i>Nardus stricta</i>	0.29	0.52	0.39	0.56	0.29	0.50	0.70	0.74
PEDVE	<i>Pedicularis verticillata</i>	0.37	0.60	0.36	0.62	0.39	0.52	0.50	0.65
PHAPI	<i>Phleum alpinum</i>	0.25	0.54	0.27	0.58	0.24	0.53	0.23	0.52
PHORB	<i>Phyteuma orbiculare</i>	0.12	0.42	0.11	0.37	0.12	0.30	0.17	0.51
PLALP	<i>Plantago alpina</i>	0.34	0.67	0.33	0.64	0.37	0.71	0.32	0.67
PLATR	<i>Plantago atrata</i> s.str.	0.28	0.55	0.44	0.71	0.30	0.58	0.28	0.58
POALP	<i>Poa alpina</i>	0.32	0.61	0.35	0.62	0.31	0.63	0.37	0.65
POAMI	<i>Poa minor</i>	0.39	0.57	0.41	0.48	0.38	0.42	0.40	0.53
POAUR	<i>Potentilla aurea</i>	0.38	0.68	0.45	0.75	0.38	0.68	0.48	0.73
POCEN	<i>Poa cenisia</i>	0.16	0.37	0.15	0.36	0.27	0.37	0.40	0.52
POCRA	<i>Potentilla crantzii</i>	0.28	0.62	0.26	0.50	0.34	0.66	0.28	0.48
POLAS	<i>Polygala alpestris</i>	0.19	0.37	0.19	0.34	0.17	0.28	0.28	0.43
POVIV	<i>Polygonum viviparum</i>	0.43	0.78	0.43	0.74	0.42	0.73	0.51	0.79
PUALP	<i>Pulsatilla alpina</i> s.str.	0.20	0.36	0.17	0.27	0.17	0.36	0.22	0.26
RAALP	<i>Ranunculus alpestris</i>	0.55	0.74	0.56	0.60	0.57	0.55	0.58	0.75
RANMO	<i>Ranunculus montanus</i>	0.30	0.58	0.32	0.58	0.29	0.49	0.47	0.73
SAAIZ	<i>Saxifraga paniculata</i>	0.17	0.46	0.17	0.37	0.18	0.22	0.15	0.18
SALRE	<i>Salix retusa</i>	0.46	0.79	0.53	0.74	0.49	0.66	0.51	0.79
SAOPP	<i>Saxifraga oppositifolia</i>	0.44	0.56	0.45	0.45	0.45	0.45	0.45	0.52
SAXAI	<i>Saxifraga aizoides</i>	0.05	0.25	0.03	0.13	0.03	0.22	0.05	0.20
SCLUC	<i>Scabiosa lucida</i>	0.20	0.47	0.17	0.43	0.19	0.36	0.25	0.56

SEATR	<i>Sedum atratum</i>	0.07	0.15	0.17	0.34	0.17	0.35	0.20	0.41
SECOE	<i>Sesleria caerulea</i>	0.36	0.68	0.49	0.69	0.35	0.66	0.43	0.69
SESEL	<i>Selaginella selaginoides</i>	0.47	0.76	0.55	0.77	0.47	0.76	0.75	0.79
SIACA	<i>Silene acaulis</i>	0.57	0.74	0.62	0.62	0.76	0.51	0.72	0.80
SILVU	<i>Silene vulgaris s.l.</i>	0.08	0.33	0.18	0.47	0.10	0.26	0.28	0.50
SOLAL	<i>Soldanella alpina</i>	0.40	0.71	0.41	0.70	0.40	0.69	0.46	0.74
TAOFF	<i>Taraxacum officinale</i>	0.19	0.44	0.26	0.44	0.23	0.47	0.17	0.26
THAPG	<i>Thymus praecox</i>	0.41	0.74	0.44	0.73	0.41	0.70	0.44	0.70
THEAL	<i>Thesium alpinum</i>	0.25	0.60	0.23	0.49	0.36	0.49	0.35	0.57
THLRE	<i>Thlaspi repens</i>	0.41	0.73	0.40	0.70	0.40	0.72	0.41	0.71
TREUR	<i>Trollius europaeus</i>	0.26	0.50	0.30	0.58	0.24	0.45	0.40	0.61
TRPRA	<i>Trifolium pratense s.str.</i>	0.27	0.61	0.35	0.65	0.30	0.54	0.47	0.73
TRTHA	<i>Trifolium thalii</i>	0.25	0.53	0.24	0.46	0.24	0.50	0.30	0.47
VAGAU	<i>Vaccinium gaultherioides</i>	0.56	0.74	0.66	0.58	0.57	0.67	0.79	0.73
VAMYR	<i>Vaccinium myrtillus</i>	0.37	0.67	0.52	0.74	0.39	0.53	0.52	0.74
VAVIT	<i>Vaccinium vitis-idaea</i>	0.39	0.46	0.44	0.49	0.43	0.47	0.69	0.61
VEAPH	<i>Veronica aphylla</i>	0.11	0.34	0.25	0.51	0.09	0.29	0.26	0.57
VICAL	<i>Viola calcarata</i>	0.29	0.61	0.30	0.54	0.38	0.68	0.38	0.61