Using predictive models to monitor and estimate population size of an invasive species: *Heracleum mantegazzianum* (Sommier & Levier)

Diploma Thesis of Biology, 2005

Christian Benetollo

Under the direction of Antoine Guisan

Table of contents

Abstract

Heracleum mantegazzianum is a potentially troublesome exotic plant species, introduced in Switzerland as a garden curiosity at the beginning of the 20th century. It forms a dense canopy out-competing native riparian species and increasing soil erosion along the stream banks where it occurs. It also represents an increasing public health hazard, because of its clear watery sap that sensitizes the skin to ultraviolet radiation. Therefore, predicting where the species is likely to spread in the future can be seen as a matter of prime importance, for: (i) directing monitoring efforts, and (ii) estimating its current population size in an area of interest. Statistical models based on a quantification of species environmental niche projected into geographical space are powerful tools in this regard, as they can serve both aims. Here, we illustrate this by presenting a predictive distribution model for *H. mantegazzianum* based on important environmental predictors and known occurrences of the species at the regional scale in the Prealps of Canton Vaud. Three predictive methods were compared: Generalized Linear Model, Generalized Additive Model and Ecological Niche Factor Analysis. A random-stratified adaptive sampling design based on a preliminary model allowed the number of individual in the region to be estimated: (i) based on the method of Thompson (*design based estimation*) and (ii) by cumulating predicted values obtained by applying the model onto the whole study area (*model based estimation*). Our results show that the species could be detected in 23 out of 100 of the squares sampled, totalizing 2590 individuals.

The species was mainly observed in disturbed habitats, like river banks and forest edges, but also along roads and railways so as in fallows and house garden, as confirmed by the selection of predictors in the models. In fact predictors which most affect the species in the study area are distance to forest edge, distance to building, slope and mean annual temperature. ENFA was the most parsimonious predictive method, probably due to the fact that it does not need absence data, and these are ambiguous in the case of an invasive species. Based on our sampling and models, the best *design-based* estimates of the population size of *H. mantegazzianum* in the whole area was 5 mios individuals (corresponding to 3% of the whole area colonisation), whereas *model-based* estimates provided values within the range of 10 to 14 mios. Hence, even the minimal estimate shows the urgency to take phyto-sanitary measures of eradication.

1. Introduction

Alien plants are species that were accidentally or deliberately introduced in a particular region since about 1500 a.C. (Jeanmonod 2002). Some of them have invasive features and causes several ecological and economical damages in the region where they have been introduced. In some cases, the species can also threaten human health (e.g. *Ambrosia artemisiifolia*, Jeanmonod 2002).

In Europe invasive species represent an important challenge on matter of conservation at the beginning of this century. The enhancements of exchanges of transports and tourism and the most accessibility to products caused by economic globalisation give rise to an enhancement in species displacement, allowing them to surmount natural geographic barriers. This also represents a significant risk of ecosystem homogenisation. Invasive alien species represent a major threat to European biological diversity with some of the rarest European species that are threatened by introduced species. It as been estimated that invasive plants, animals and insects are the second-largest threat to biodiversity, after habitat degradation (Hoag, 2003).

Most of alien plants are not dominant competitors in their native ecosystem and some of them have not an invasive pattern. There are several hypotheses as to why some species increase uncontrollably and become pests when they are introduced to new area. The most frequent explanation is that parasites and pathogens of an introduced species are often left behind their hosts when they migrate, well known as “natural enemies release hypothesis” (Mitchell & Power 2003, Torchin *et al.* 2003, Clay 2003). Furthermore, the absence of natural enemies in the introduced area may allow alien species to invest less in defence, and consequently to use more energy for other traits of life, like growth, fecundity and competition (Evolution of Increased Competitive Ability (EICA) hypothesis) (Blossey & Nötzold 1995). Callaway (2004) suggest that invasive plants increase competitive ability because they possess novel “biochemical weapons” that function as usually powerful allelophatic agents, or as mediators of new plant-soil microbial interaction.
Beside, there is strong evidence that introduced plants frequently hybridize with native relative and with other introduced taxa (Abbot 1992, Rhymer & Simberloff 1996, Ellstarnd & Schierenbeck 2000, Vilà et al. 2000, Lambrinos 2004). In a dynamic of invasion a very important parameters is local adaptation. There is growing evidences from a range of taxa that introduced populations can quickly adapt to local conditions (Reznick and Ghalambor 2001, Grosholz 2002, Lee 2002) supported, most in the early stages of invasion, by an elevated phenotypic plasticity (Baker 1974,

Sexton *et al*. 2002). This quick adaptation might be related to specific species traits (e.g. polyploidism) and favoured by hybridisation.

It seems also clear that certain habitats are more prone to invasion than others (Clay 2003). Elton's hypothesis (1958), suggests that an important determinant of invasion success is resident biodiversity, arguing that high diversity increases the competitive environment of communities and makes them more difficult to invade. The work of Sahid *et al.* (2000) show that the expectation of Elton was true, but some covarying extrinsic factors may obscure the negative impact of diversity on invader success. Such covarying factors include disturbance, fire frequency, grazing, proximity to human population, and some others.

Indeed disturbed habitats are known to be very favourable to alien plants because only pioneer native species can live there, hence reducing competition and favouring spread of propagules. Consequently, invasive plants can often be observed, along roadsides, river banks, railway embankments, cemeteries and fallows, most of the time in the early stages of invasion. In particular, the effect of fire can be very supportive for the success of invasive plants, as reported by Keeley et al. (2003) in their study in southern Sierra Nevada, and by Keeley (2002) in California.

In Switzerland according to the Red Lists (Moser *et al.* 2002) the flora presently encompasses a total of 2943 species of which 350 (12%) are alien species, i.e. species naturalised in Switzerland after 1500 due to human activity. Only 10% of these can cause problems. As negative effects have not yet been observed for the remaining 90%, these alien species can be viewed as an enrichment to the flora (Gigon & Weber 2004).

In Switzerland some organisations have been created to control invasive plants and find solutions to the problems they are causing. One of these is the Swiss Commission for Wild Plant Conservation (CPS-SKEW), created in 2001, which task had been to classify plant species into different lists according to the degree of danger to the environment, the economy and the human health (Gigon & Weber 2004). The Black List is the register of alien invasive plant species in Switzerland that actually cause damage. The Watch List is the register of alien invasive plant species of Switzerland which have the potential to cause damage or already cause damage in the neighbouring countries and/or are on an official Black List or similar in this countries. The dispersal and the effects of these species should be regularly documented. If necessary, control measures should be taken (Appendix F).

Damages caused by invasive plants to ecosystem and economy begin to become a very important problem in Switzerland, so as in most countries over the world. Trading and distribution of these plants should therefore be controlled and prevented, and after a cost benefit analysis, a professional, focused and efficient control should be started, especially for species of the Black List. This is a requirement of the Convention on Biological Diversity of Rio that was also ratified by Switzerland (Gigon & Weber 2004). However to undertake an efficient control process it is important to focus attention on those areas that are more prone to be invaded by exotic species. In this regard, predictive modelling of species distribution could prove very useful.

The first aim of the present study is thus to determine the potential distribution of a problematic invasive species, *Heracleum mantegazzianum* using methods of geographical information system (GIS) and statistical methods like Generalized Linear Models (GLM, McCullagh & Nelder, 1989), Generalized Additive Models (GAM, Hastie & Tibshirani, 1986) and the Ecological Niche Factor Analysis (ENFA, Hirzel *et al*., 2002). Such predictive models were also used in other works concerning invasive species (Peterson *et al.* (2003), Pysek *et al*. (1998), Joshi *et al.* (2004)). A series of climatic and other ecogeographical predictors are compared with presence/absence data and the corresponding abundances of the species, to understand which factors are important to determine its presence. A second aim was to estimate the total number of individuals in the whole study area using design-based and model-based techniques of estimation. These could be very important tools to monitoring and manage an invasive species.

2. Methods

2.1 Heracleum mantegazzianum

Heracleum mantegazzianum Somm. et Levier (*Apiaceae*) is a perennial monocarpic herb with a thick taproot, a stout stem attaining 5.5 m height, large pinnate leaves and usually seven to ten umbels bearing oval-elliptical, broadly winged fruits of 9-11 mm in size (Tiley *et al.* 1996). The species spreads exclusively by seed, and the seed production per plant may reach several tens of thousands (Pysek et al., 1995). Germination occurs in most central Europe in early spring (April), flowering period starts in May, fruits appear in July and are shed from September onwards (Tiley *et al*. 1996).

The species is native to the Caucasus, south-west Asia, where it occurs in forest edges and glades, often at stream sides in montane areas with annual rainfalls between 1000 and 2000 mm (Pysek 1991; Ochsmann 1992). Here the climate is typically continental, with hot summer and cold winters, tempered somewhat by the forest and upland environment, giving humid condition (Mandenova 1950). It was introduced in central Europe in the earlier 19th century mostly in botanical gardens, as a curiosity. The invasive character of the Giant Hogweed is particular due to its high production of seeds. Dispersion can occur locally around the stalk, but long-distance dispersal occurs naturally mostly along water courses: flotation tests indicate that the fruits float in water for up to three days (Clegg & Grace 1974; Dawe & White 1979). Secondary spread readily occurs onto roadsides, railway embankments and cemeteries. Human activity such as the collection of fruiting heads also contributes to spread (Lundström 1984). Another part of the species' competitive superiority over other plants is ascribed to its size and its ability to shade the surrounding vegetation with its huge ground leaves. In winter it increases soil erosion along stream banks where it occurs and it enhances the quantity of nutritive matter in the rivers causing serious damages to the local ecosystems. Not only ecological dangers are attributable to this plant. Indeed in Switzerland it also represents an increasing health hazard because of its clear watery sap that sensitizes the skin to ultraviolet radiation (Tiley, 1996). However it has been demonstrate that cattle, sheep and goats frequently browse the plant with impunity and no apparent ill effects was reported from countries where *H. mantegazzianum* is used for silage.

2.2 Study area

The study area covers an area of about 564 square kilometers, encompassing nearly all mountains of the PreAlps of Canton de Vaud (Fig. 2.1). Elevation ranges from about 400m in the lowland Rhone Valley up to more than 3000 m at the top of the highest peak of the area (Diablerets 3210m) where important glaciers are still found. The climate of the region is essentially wet and fresh. The rainfalls are abundant and grow with increasing elevation. The temperature is fresh during the summer and can be very cold during the winter, with important precipitation and snowfalls. Vegetation has been and still is very influenced by human activity: the pasture is common in this region and is observed from the valley bottoms to the subalpine and lowers alpine areas.

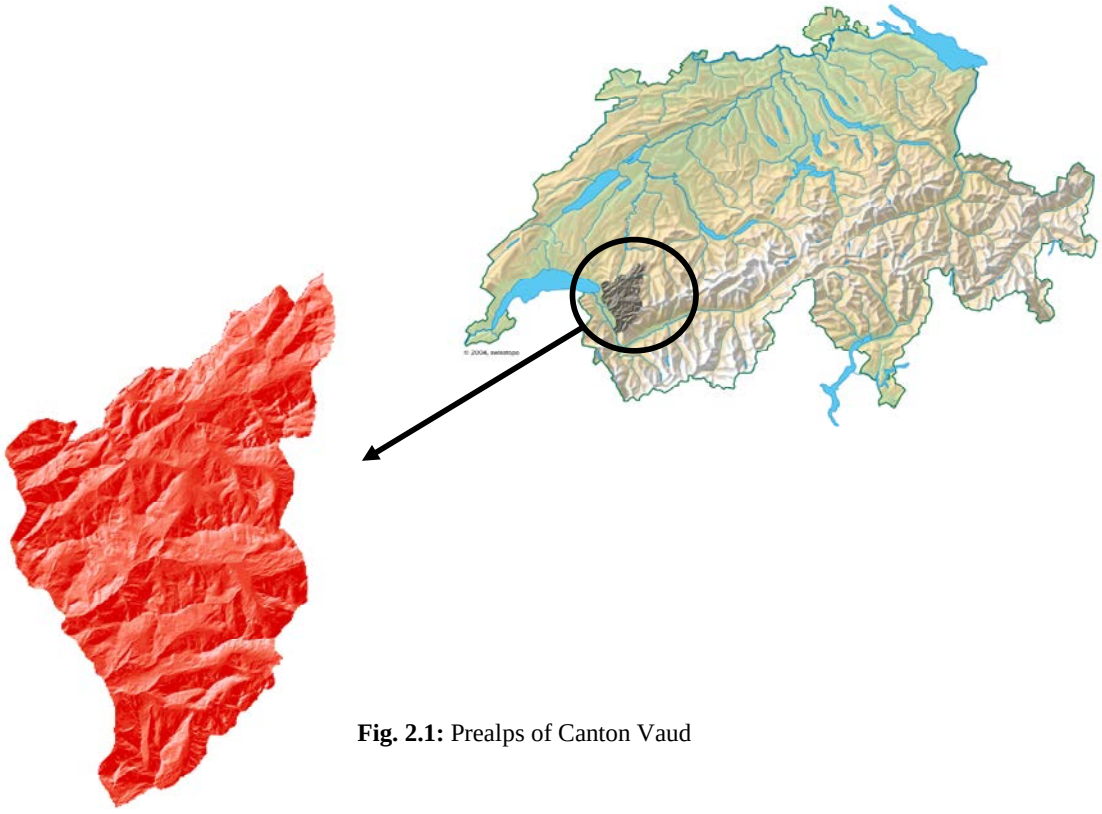

Fig. 2.1: Prealps of Canton Vaud

2.3 Environmental predictors and initial species data

For each analysis ecogeographical predictors were selected. These predictors can be separated in quantitative, semi-quantitative or qualitative (Table 2.1).

Table 2.1: Quantitative, semi-quantitative and qualitative predictors used in the analysis

Name	Abbreviation	Unity	Category
elevation	mnt	m	quantitative
annual sum precipitation	precyy	mm	quantitative
number of day of frost	sfroyy	day*100	quantitative
slope	slp	deg	quantitative
solar radiation	sradyy	KJ / day	quantitative
annual average site water balance	swb	1/10mm/yr	quantitative
annual average temperature	taveyy	°C*100	quantitative
topographic position	topo	unitless	quantitative
topographic wetness index	twi	unitless	quantitative
road distance	roaddist	unitless/m	semi-quantitative/quantitative
river distance	riverdist	unitless/m	semi-quantitative/quantitative
forest edges distance	foresteddist	unitless/m	semi-quantitative/quantitative
building distance	builddist	unitless/m	semi-quantitative/quantitative
buffered roads	broads	-	qualitative
buffered rivers	brivers	-	qualitative
buffered railways	brailways	-	qualitative
buffered forest edges	bforested	-	qualitative

Topographic position takes negative values for concavity positions and positive values for convexity positions. For semi-quantitative distances 6 ordered classes were created. Class 0: 0m, Class 1: 0-25m, class 2: 25-50m, class 3: 50-100m, class 4: 100-200m, class 5: >200m.

Precyy, *sfroyy*, *slp*, *sradyy*, *swb*, *taveyy*, *topo*, and *twi* (table 2.1) were calculated following calculations of Zimmermann and Kienast (1999).

Quantitative predictor's distances were calculated with a neighbourhood statistic in the ArcMap software (Esri ArcMap 8.2, 1999-2002 Esri Inc) with a sum statistic type. Qualitative parameters were obtained by considering a buffered zone around those objects that could influence species occurrence. A buffer of 15 m was applied to roads and railways and a buffered zone of 25 m was considered for rivers and forest edges, considering the fact that this plants disperse and spread essentially along communication ways and rivers.

2.4 Adaptive sampling

A random stratified adaptive sampling was applied to the entire area. Ten strata corresponding to ten probability classes from a preliminary habitat suitability map for *H. mantegazzianum* in Switzerland (Benetollo & Monico, 2003) were used. In this work ENFA habitat suitability map was calibrate at Swiss scale with 92 presence data obtained from the «Centre du Réseau Suisse de Floristique» (CRSF) and some ecogeographical parameters (annual average temperature, annual sum of precipitation, topographic wetness index, annual average site water balance, solar radiation, slope and continentality index). The map was finally filtered with some qualitative variable (roads, rivers, forest, and landuse) which can enhance or decrease the probability of the species installation.

Ten 25m x 25m squares were randomly sampled in each stratum. Adaptive sampling (Thompson & Seber, 1996) then dictates that, for each sampling cell where the species is observed, a reiteration of the sampling for the four neighboring squares is conducted. If the species is once more observed in any of the neighboring cells the same procedure is repeated in the adjacent cells, and so on until no more occurrence is found in neighboring cells (Fig.2.2). This process is particularly adapted in the case of rare and patchy distributed species like the Giant Hogweed. For each presence cell, an evaluation of the number of individual and a description of the habitat was also made.

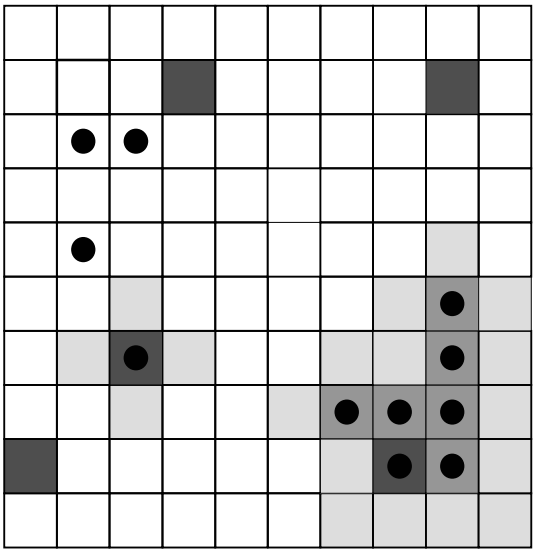

Fig. 2.2: Random adaptive sampling. Five squares are sampled randomly (dark grey). Only in two of them the species is present (black points). In this case a reiteration of the sampling occurs: six new squares are sampled (grey). Hell grey squares represent sure absences.

2.5 Modelling methods

Habitat suitability maps were obtained using two different methods. Initially only presence data were available and thus an Ecological Niche Factor Analysis (ENFA) was used. When absences data became additionally available Generalized Linear Models (GLM) and Generalized Adaptive Models (GAM) can be used instead. The latter usually result in greater predictive power. Correlations between quantitative predictors also need to be investigated, as they can affect model fitting. When two or more predictors had a correlation coefficient greater than 0.8, only the most causal was kept (Appendix A, table 1)

2.5.1 ENFA

The Ecological Niche Factor Analysis (Hirzel *et al*., 2002) is a method based on the comparison between the environmental niche of the species and the environmental characteristics of the entire study area. It is a multivariate approach to the study of geographic species distribution developed for cases where only presence are available. ENFA is very similar to a kind of principal component analysis (PCA). It also transforms the original ecogeographical predictors into new uncorrelated axes (factors).

The first factor is extracted in a way that maximizes the marginality of the focal species, defined as the ecological distance in multidimensional environmental space between the species optimum and the mean habitat within the reference area (i.e. the axes passes through the two centroïde). If the marginality value associate to a predictor is negative, it means that species optimum value is lower than the mean value of the study area. The other factors maximizes the specialisation of this focal species, defined as the ratio of the ecological variance in mean habitat to that observed for the focal species. For each predictor the specialisation value has to be considered as an absolute value. Eigenvectors and eigenvalues are readily interpreted and can be used to build habitat-suitability maps. The only disadvantage is that the predictive variables must follow a normal distribution. Our analyses were performed with the Biomapper software (version 2.1, Hirzel 1998 - 2002). The broken stick method (Frontier, 1976) was used to select the number of principle components. These were created from series of quantitative predictors, first normalized through the Box-Cox algorithm (Sokal & Rohlf 1981), using only 75% of the presence data. The rest was used to validate the model and finally tested by a bootstrap method applied on the evaluation data set.

2.5.2 GLM and GAM

Generalized Linear Model (GLM, McCullagh & Nelder, 1989) is an extension of ordinary least square regression, allowing non-normal response variables to be modelled. Our GLM analyse were performed in the R software (R 1.9.1; A Language for Data Analysis and Graphics © 2004). Important predictors were selected by a stepwise selection procedure. Generalized Additive Model (GAM, Hastie & Tibshirani, 1986) is an extension of GLM where the linear form is replaced by a sum of smooth functions. GAM analyses were run in GRASP 2.5 (Generalized Regression Analysis and Spatial Prediction, Lehmann *et al*. 2002). Four smoothing degrees of freedom were allowed and a stepwise AIC was use to select important predictors.

For presence/absence response variables, a binomial family was used. For individual counts, a binomial GLM and GAM with logit link were used as well, as in our case these are bound by a minimum (floor) value of 0 and a maximum number of plants per cell of 200 (ceiling). Indeed, an estimation of the maximal number of individual that can occupy a square of 25m X 25m was done. To do this we estimated the area covered on the ground by an adult individual. We considered a circle area with 1 meter radius as leaves can reach a length of 1 m (Jeanmonod 1999), resulting in a maximal number of individual per square of 200. Counts were divided by this maximum abundance value to get values of relative abundance (a relative measure of success) that could be modelled as probabilities using a binomial GLM or GAM.

Potential distribution maps were built in the scale of the linear predictor based on the coefficient estimates for each selected predictor. The logit link function was used to obtain values in the scale of the response variable, namely the probability of presence.

2.5.3 Evaluation of models

GLM models were evaluated by different methods. First in diagnostic plots obtained in R residuals were plotted against predicted values (in the scale of linear predictors). Such a plot is useful for looking for outliers, non-linearity, and to check the assumption of constant variance. The largest residuals are also identified. A second plot shows ordered deviance residuals plotted against normal order statistics (Davison, 1989).

An evaluation data set has been created with a 10 fold cross validation procedure (Guisan & Zimmermann, 2000). The presence/absence threshold maximizing *Kappa* value was first calculated: the species is considered absent in pixels with a probability lower than the threshold, and present elsewhere. The contingency table determining the kappa value for this threshold was obtained for

both calibration and evaluation data set. The ROC-plot statistics, which does not require the calibration of a threshold, has been carried-out on the calibration and the evaluation data set. The correction of Gini (Hand & Henley 1997, Copas 1999) was applied to the area under the curve (AUC) of a ROC plot to obtain value ranging between 0 and 1 (AUC'). Otherwise jack-knife with his corresponding kappa value, and a bootstrap (Efron & Tibshirani, 1993) have been performed. For GAM models, ROC-plot test was applied on the calibration data set, and on the estimation data set obtained by cross-validation on 10 groups. In each case an uncertainty map is obtained subtracting the map constituted from the maximum plot values of every predictor (obtained by summing the estimate value of each predictor with the standard deviation) and those characterised from the minimum plot values (obtained by subtracting the estimate value of each predictor with the standard deviation).

2.5.4 Comparison between habitat suitability maps

Two compare the three habitat suitability maps we calculate the Kappa value between two maps considering only two classes of presence probability (0 for presence probability between 0 and 0.5; 1 for presence probability between 0.5 and 1). If all pixels with value 1 in the first map are the same than in the second map Kappa is 1.
Another way to compare habitat suitability map is to calculate a congruence map between two maps. An high congruence value means that the pixel has the same presence probability value for the two maps.

2.6 Estimation of species density in the study area

To estimate the number of individual of *H. mantegazzianum* in the study area we compared two different conceptual approaches: *design based estimation* and *model based estimation*.

2.6.1 Design based estimation

Thompson & Seber (1996) suggest the use of an adaptive sampling to estimate density for rare and clustered species. Each presence point generates a network composed by itself (dark grey squares in fig. 2.2) and by the reiterated presence points (grey squares). Hence a network is composed from 1 to *u* units (square 25 X 25m), and each units contains a certain number of individuals (y-value)

(Thompson & Seber, pp.113-117). In their method the intersection probabilities for each network must first be calculated (the method also considers cases where the network is composed by two or more strata):

$$\alpha_i = 1 - \left[\binom{N_s - m_i}{n} \Big/ \binom{N_s}{n} \right] \quad (1)$$

Where N_s is the total number of unit in a strata *s*, *n* is the number of sampled unit for each strata and m_i is the number of units in a network *i*.
The mean number is estimated as:

$$\mu^*_s = \frac{1}{N_s} \left(\sum_{i=1}^{i} \frac{y_i}{\alpha_i} \right) \quad (2)$$

Where y_i is the total number of individuals in network i. The estimated variance is:

$$\text{var}[\mu^*_s] = \frac{1}{N^2} \left[\frac{{y_1}^2}{\alpha_1} \left(\frac{1}{\alpha_1} - 1 \right) + \frac{{y_2}^2}{\alpha_2} \left(\frac{1}{\alpha_2} - 1 \right) + \frac{2 y_1 y_2}{\alpha_{12}} \left(\frac{\alpha_{12}}{\alpha_1 \alpha_2} - 1 \right) \right] \quad (3)$$

Where $\alpha_1 \alpha_2$ is calculated as follow:

$$\alpha_{12} = \alpha_1 + \alpha_2 - \left(1 - \left[\binom{N - (m_i + m_{i+1})}{n} \Big/ \binom{N}{n} \right] \right) \quad (4)$$

2.6.2 Model based estimation

The second way to estimate the number of individuals in the study area is to cumulate predicted values obtained by applying the model to the whole study area (model based estimation). This

estimation is highly dependent upon the model precision. The predictive methods employed for this estimation were GLM and GAM with a binomial function. In this case we used the proportion of individuals on 200 as response variable (see modelling methods). Probabilities were then multiplied by 200 to obtain a map in the scale of the response (i.e. number of individuals), which was finally filtered by forest, building and over 1800 m altitude layers.

3. Results

3.1 Field sampling

H. mantegazzianum was present in 23 of the 100 sites sampled. Networks were composed of one to eight units, and interestingly the maximal number of individuals in a square turned out to be exactly 200 (Fig. 3.1) i.e. our theoretical maximum (see method section). If we also consider reiterated units, 53 presence cells and 219 absence cells were visited. In total 2590 individual were observed within an altitudinal range between 612 and 1620 meters. Individuals were found at different stages of their life cycle. Sometimes only leaves were detectable. In these cases, species determination was difficult because of the similarity in earlier life stages between *H. mantegazzianum* and *H sphondylium*, the native closest relative species. The most developed population were found in Les Diablerets (1309) and Villard (1480 m). No individuals were observed in the plain. The species was found in different habitats, mostly on river banks and forest edges, but also along roads and railways, so as in fallows and gardens. Some individuals were even found in forests where canopy was not too dense; on the contrary no individual was detected in grasslands.

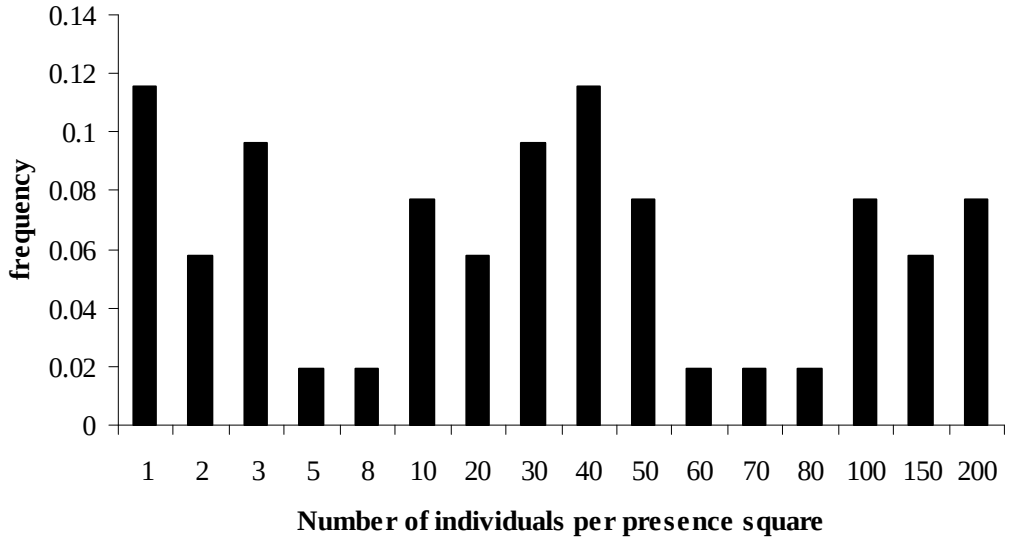

Fig. 3.1: Histogram showing the distribution of the number of individuals per cell; y axis is in number of cells

3.2 Predictive models

3.2.1 Ecological Niche Factor Analysis

The global coefficient of marginality was 0.635, the overall tolerance was 0.161 and specialisation (1/tolerance) was 6.223. The brocken-stick method selected the first four factors (Table 3.1) explaining altogether 94.5% of total variance.

Table 3.1: Eigenvalues for the four first factors explained variance and cumulate explained variance

Factor	Value	Expl.Var.	Cum.Expl.Var
1	242.791	0.627	0.627
2	65.279	0.169	0.796
3	35.009	0.09	0.886
4	22.911	0.059	0.945

The ten environmental predictors explained different portions of the first two ENFA factors (Table 3.2).

Table 3.2: Coefficient of the ecogeographical predictors for the first four factors

Marginality (63%)		**Specialisation 1** (17%)		**Specialisation 2** (9%)		**Specialisation 3** (6%)	
builddist	-0.45	roaddist	0.71	roaddist	-0.56	taveyy	-0.77
riverdist	-0.40	builddist	-0.52	foresteddist	0.56	builddist	-0.42
slp	-0.39	foresteddist	0.38	taveyy	-0.48	swb	-0.32
sradyy	0.39	topo	-0.20	slp	-0.24	sradyy	-0.26
topo	-0.31	taveyy	0.16	swb	-0.20	foresteddist	-0.19
foresteddist	-0.30	riverdist	0.11	topo	-0.12	riverdist	0.11
roaddist	-0.25	sradyy	0.09	builddist	0.12	topo	-0.11
swb	0.23	twi	0.06	twi	-0.12	slp	0.07
twi	0.18	slp	0.04	riverdist	-0.07	twi	0.00
taveyy	0.04	swb	0.03	sradyy	0.04	roaddist	0.00

In the Prealps of Canton Vaud, *H. mantegazzianum* prefers habitats located near buildings (marginality value = -0.45) and rivers (marginality value = -0.4) with a weak slope (marginality value = -0.39) and high solar radiation (marginality value = 0.39). The preferred temperature is close to the mean temperature (marginality value = 0.04) of the study area. Furthermore table 3.2 shows that the species is very little tolerant to greater building distances (specialisation 1 value = -0.52) so as to the road distance (specialisation 1 value = 0.71, specialisation 2 value = -0.56); on the contrary it is very tolerant to slope (specialisation 1 value = 0.04) and solar radiation (specialisation 1 value = -0.09, specialisation 2 = 0.04). Figure 3.5 (a) shows the resulting habitat suitability map.

The mean presence probability of evaluation data set was 69.4%, whereas it was only 39.2% for the calibration cells (Appendix B, Fig. 1). 71.4% of validation cells had a habitat suitability value over 50% (bootstrap test = 0).

3.2.2 GLM models

The stepwise procedure retained four predictors: distance to forest edge (semi-quantitative parameter), slope, mean annual temperature and topographic position (Table 3.3; Appendix C, table 1 and 2). The model explained 36% (D^2) of the total deviance, with an adjusted explained deviance of 30% (adj.D^2).

Table 3.3: parameters retained in GLM with their respective significance

	Estimate	**Std.Error**	**Significance**
(Intercept)	-12.00	5.00	*
foresteddist	2.44	$7.53*10^{-1}$	**
$(foresteddist)^2$	$-4.22*10^{-1}$	$1.49*10^{-1}$	**
slp	$2.86*10^{-1}$	$1.53*10^{-1}$	.
$(slp)^2$	$-8.49*10^{-3}$	$4.04*10^{-3}$	*
taveyy	$3.16*10^{-2}$	$1.58*10^{-2}$	*
$(taveyy)_2$	$-2.72*10^{-5}$	$1.30*10^{-5}$	*
topo	$1.02*10^{-2}$	$4.28*10^{-3}$	*

H mantegazzianum's response to distance to forest edge, temperature, precipitation and slope was best fitted with a unimodal response curve. A distance of 50 – 100 m to edge forest, a slope of 15° and a mean annual temperature of about 5°C seems to be optimal habitat conditions for the species (Fig. 3.2). On the contrary, Giant hogweed has a linear response to topographic position (Fig. 3.2).

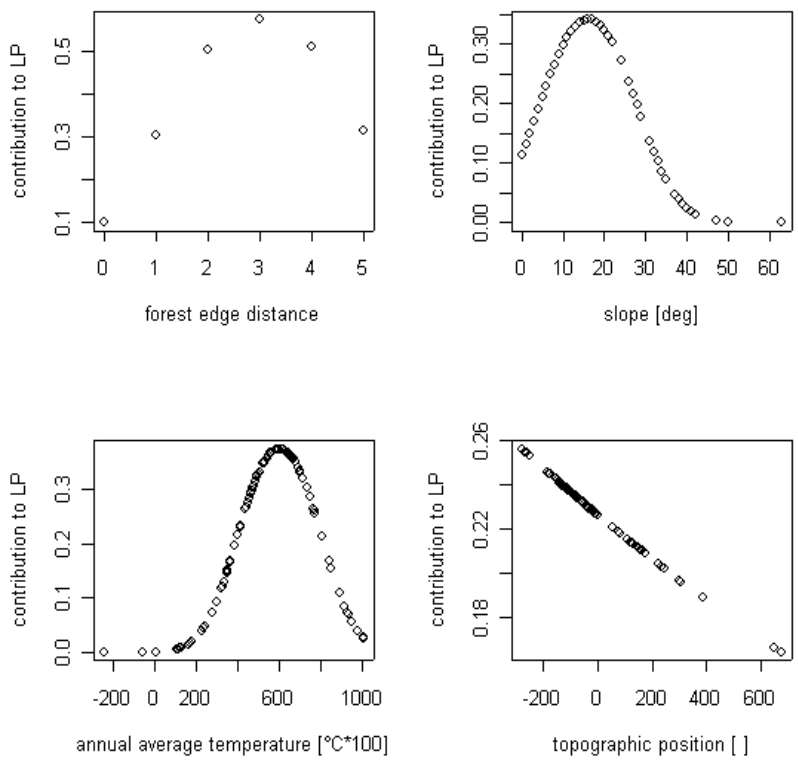

Fig. 3.2: GLM response curves for the four predictors

The resulting habitat suitability map (Fig. 3.5 (b)) shows in red pixels with the highest probability of presence of the species, and in green the lowest probability. The uncertainty map shows that greatest error values correspond to high probability of presence.

Figure 3.3 a) shows residuals plotted against predicted values: as values range between values of -2 and 2, the assumption of constant variance is confirmed. The three largest residuals are also identified (data 49, 61 and 97). Data 49 and 97 corresponds to presence plots, data 61 to an absence plot. Ordered deviance residuals plotted against normal order statistics shows that the distribution is positively skewed relative to the normal distribution (Fig.3.3 (b)), which means that residuals follow a normal distribution.

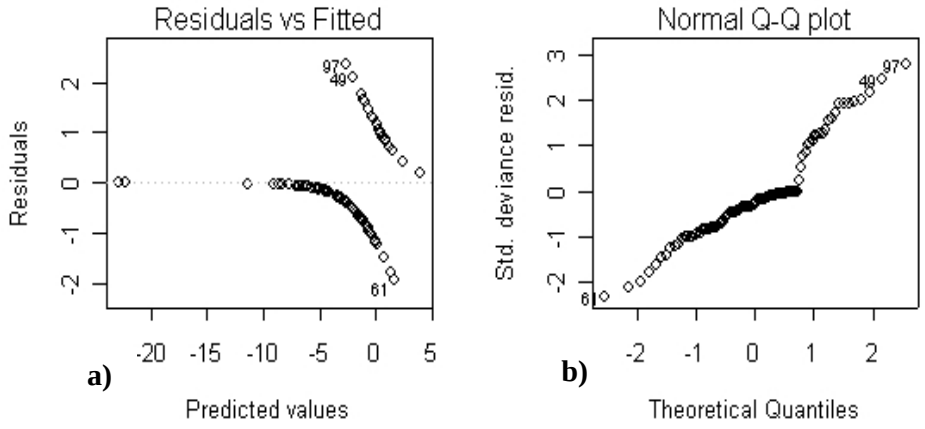

Fig. 3.3: Diagnostic plots GLM (predicted values are represented in the linear predictor scale)

The threshold value maximising kappa was 0.55. The kappa value for this threshold is 0.54 for calibration and 0.35 for evaluation data set. A lower value for the evaluation set was expected because it did not consider some situations absent from the training set. The contingency table for this threshold shows that 72/77 of absence and 13/23 of presence were correctly predicted (Table 3.4).

Table 3.4: Contingency table of GLM (a = absences, p = presences)

		predicted	**values**
		a	p
observed	a	72	5
values	p	10	13

The ROC approach gives an AUC' value of 0.77 for the calibration set and 0.63 for the evaluation set.

Kappa value of jackknife validation method was 0.4, and the bootstrap evaluation showed that the difference between estimated predictor values and the real population values (bias) was $< 1* 10^{-15}$.

Uncertainty map (Fig. 3.5 (d)) shows that high altitude areas are associated with small uncertainty. This is probably because temperature was retained in the model, and only absences were detected for low values of this predictor (i.e. high values of elevation).

3.2.3 GAM models

GAM retained three predictors, explaining 38% of the total deviance: distance to forest edge (quantitative parameter), slope and topographic position. Response curves and error curves for each predictor are presented in figure 3.4.

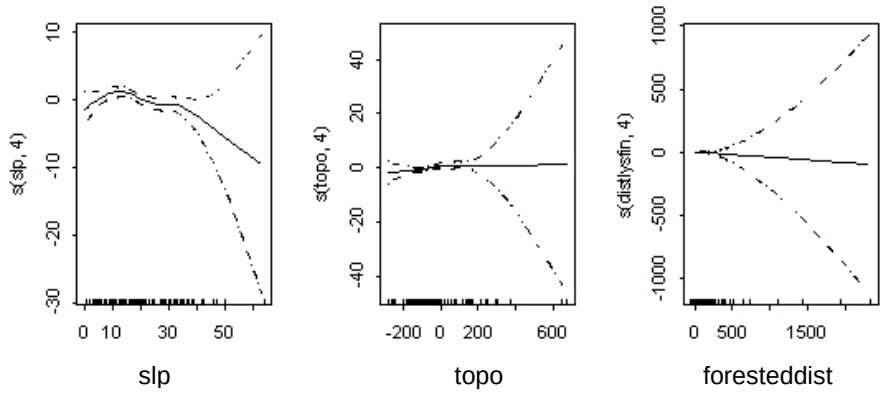

Fig. 3.4: GAM response curves for the three predictors

For the three predictors we obtained the biggest error for the highest value, which in each case are associated to high elevated zone (big slope, convex topographic position and the highest distance to forest edge). Furthermore distribution of sampled point is not homogeneous along gradients: for the distance to forest edge there are only four points over 1000m!

If we compare the response curve of GAM with those of GLM we see that slope have also a unimodal response with the optimum at the same value (15°). Surprisingly topographic position has an opposite response in front of GLM, actually in GAM species seems prefer concave position. Finally in GAM, distance to forest edge has a linear response with the highest value for low distances whereas in GLM the response was unimodal.

Once again the resulting habitat suitability map (Fig. 3.5 (c)) shows in red pixels with the highest probability of presence of the species, and in green the lowest presence probability.

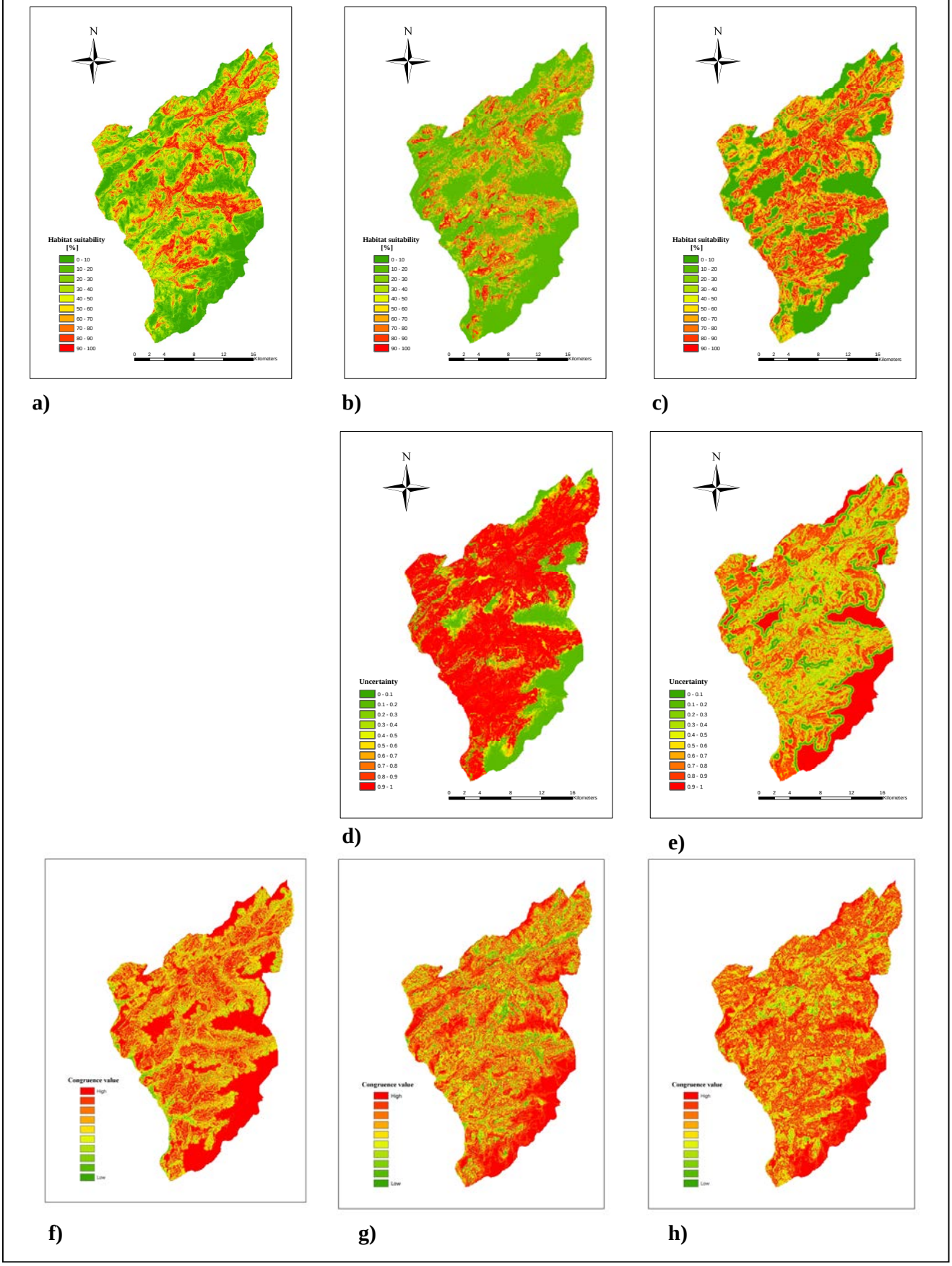

Fig. 3.5: ENFA (a), GLM (b) and GAM (c) habitat suitability map. GLM (d) and GAM(e) uncertainty map. Congruence map between GLM and GAM (f), GLM and ENFA (g), GAM and ENFA (h).

The evaluation with the ROC approach gives an AUC value of 0.89 and a value for Gini (AUC') of 0.76. The AUC obtained by cross-validation is 0.76 and 0.52 for CV-AUC'.

The uncertainty map (Fig. 3.5 (e)) shows this time the inverse pattern than obtained for the GLM, namely high elevation areas present the highest uncertainty. Indeed, the three predictors retained by the model have the highest standard deviations for values corresponding to high elevation zone (Fig. 3.4).

To compare the similarity between maps we calculated Kappa values:

GLM-GAM: Kappa = 0.738 (error: 0.26)
GLM-ENFA: Kappa = 0.666 (error: 0.33)
ENFA-GAM: Kappa = 0.654 (error: 0.34)

These results associated with congruence maps of figures 3.5 (f,g and h) which show in red pixels with the same presence probability, suggest that GLM and GAM are the most similar model. This is not surprising given that this two predictive model are both based on presence-absence data.

3.3 Estimation of the total population size in the study area

i)

Based on Thompson's calculations the best design-based estimates are:

$$\mu^* = 5\ 259\ 924.25 \quad (\text{var} = \quad 191\ 992.7)$$

To better understand the meaning of these values we estimated the population size corresponding to a complete, although improbable invasion of the total study area by the species considering that the maximal number of individual per plot is 200 (see chapter 2). This maximum estimate was:

$$\mu_{tot} = 180\ 352\ 000$$

Hence our previous estimate (μ^*) corresponds to about 3% of the complete invasion.

ii)

The binomial model, obtained with a GLM and a stepwise procedure retained four parameters: slope, topographic position, distance to forest edge (quantitative parameter) and annual sum of precipitation (Table 1 and 2, Appendix E). This last one was taken in account in stead of the annual average temperature because most predictive. This model explains 49% of the total deviance (D^2). The GAM model retained three predictors: forest edge distance, slope and annual sum of precipitation (Fig.1, Appendix E), in this case the explained variance was 64%.
Based on cumulating the predicted number of individuals in each pixel of 25m x25m obtained from the binomial model we obtain the following results:

$$\mu_{GLM} = 14\ 574\ 714$$

and

$$\mu_{GAM} = 10\ 781\ 419$$

These estimations correspond at 8% and 6% of the complete invasion.

3.4 Comparison of ENFA model calibrate at global and local level

Here we report the comparison of an ENFA model calibrated at Swiss scale (Benetollo & Monico 2003) and at the scale of the Prealps of Canton Vaud (Fig. 1, Appendix D). Model at Swiss scale was fitted at a coarser resolution with a pixel grain of 100m X 100m. The proportion of variance explained was 81.7% whereas it reached 97% for the local model. Even the marginality and specialisation values were different: 0.586 and 0.635 for the marginality; 0.456 and 0.161 for the specialisation. Data were also collected in different period: the work of Benetollo & Monico was based on observation from 1948 to 2002, whereas the local model was based on observations made during this year.

4. Discussion

Heracleum mantegazzianum was voluntarily introduced in Switzerland in some botanical gardens but also in some private gardens in villages like Villard and Les Diablerets, which are typical Swiss vacation villages. Hence, it is not a surprise to find the largest populations there. Furthermore, owners of hotels or villas still think that such a plant can beautify their garden, ignoring troubles that it can provoke. Escaping from these sources the species spreads secondarily along riverbank and roadsides, reaching other favourable habitats, typically in disturbed habitat like forest edges, glades, fallows and abandoned farmlands. These considerations match very well with the work of Pysek (1994) in the Czech Republic.

4.1 Predictor responses

Distance to forest edge was retained as an important predictor for the species' presence: ENFA results show that *H. mantegazzianum* is mostly found near forest edges and that it is quite specific to low distances to forest. Additional information is given by GLM analysis: the preferred distance to forest edge is 50 – 100 m. The fact that the species grows and reach its adult stage preferably in forest edges is because it is continuously erased by mowing and browsing activity in adjacent grasslands and pastures, although these could be favourable to the species.

Although rivers represent the most important vector of dispersal for the species, they do not seem to be an important predictor for determining its presence. The ENFA analysis shows that *H. mantegazzianum* prefers habitats near rivers, but that it also tolerates other conditions further apart. This could be an evidence of an advanced expansion phase; where the species exits from typical spread corridors and invades secondary habitats. Actually dispersion probably occurs mostly along roads and railways. The work of Pysek & Prach (1993) shows the same pattern. They found that, compared to other riparian invasive species, *H. mantegazzianum* has a lower proportion of riparian localities, probably due to its ability to expand massively to other habitat types. This fact may be explained by its high competitive success over native species in a wide range of habitats (Pysek 1993). Furthermore, animals and wind also play an important role in its dispersion (Jehlik & Lhotska 1970; Williamson & Forbes 1982; Lundström 1984; Wyse & Jackson 1989; Pysek 1991).

In the ENFA analysis the effect of distance to building is very strong. We can find the species very close to buildings with a pretty high value of specialisation. Distance to buildings can be considered as a surrogate for human density that suggests a close relationship between distribution of *Heracleum mantegazzianum* and density of human population. This reflects that intensity of human disturbances provides safe sites for establishment and suitable habitats for subsequent invasion. This is also in accordance with the work of Pysek *et al*. (1998) who found that density of human population and January isotherm were the most important predictors determining its presence in the Czech Republic. In our study, temperature plays an important role as well, although mostly in GLM. *H. mantegazzianum* is more represented in regions with an annual average temperature of 5 °C. Similarly, Pysek (1998) found that regions most suitable to the species have a January isotherm of 3°C. Mean annual temperature marginality value and specialisation value suggest that the species finds optimal temperature conditions for growth in the study area.

Slope, and particularly gentle slope, is also an important determinant of the presence of this species in particular small values of slope are more appreciated by the species. The fact that ENFA specialisation value for slope is very low suggests that the species can also be present on steeper slopes, where habitats are more disturbed. Once again, our results confirm the works of Pysek (1998). Results for topographic position suggest that *Heracleum mantegazzianum* is more frequently found on valley bottoms than on exposed area. This is not surprising as the species was frequently found along rivers. Surprisingly for this predictor, GAM response curve had the inverse pattern than GLM, but in the first case highest values of topographic position are also associated to elevate standard errors (fig. 3.4).

In general, species is marginal to conditions of the study area but it tolerates very well other conditions. This is very characteristic of an invasive species which can colonise different habitats.

It has to be noted that no qualitative parameter was retained by the predictive models probably because buffered zones were too narrow, and predictive models could not discern presences from absences.

Patterns discussed above can be visible also in the resulting habitat suitability map. In general, the habitat suitability maps for GLM, GAM and ENFA show the same pattern (fig. 3.5 (a), (b) and (c)). Giant hogweed is mostly predicted along valleys below 1800 meters. GAM presents a higher proportion of pixel with presence probability between 90% and 100%; on the contrary ENFA has a more balanced proportion of each presence probability class.

4.2 Comparison between predictive models

To compare the three predictive methods, we first analyse what were the most important differences between ENFA and the other two methods. In fact, it is important to understand how much absence data are important for predicting the distribution of an invasive species. Invasive species often do not have reached their equilibrium situation, which means that some absence data observed in the field are not true absence. Species did not yet have the time to colonise these regions, for historical reasons or for dispersal limitations. The risk is to have a larger proportion of absence pixels with the same predictor value than presence pixels, and the predictive method cannot then discriminate between the two different classes. For these reasons, ENFA, which requires only presence data, is the most parsimonious predictive method. In fact, if we validate the model with absence data verified in the field, we see that there are 40 out of 77 absence plots that have a presence probability between 75% and 100%. There are two reasons to explain the fact that the GLM and the GAM respectively explain only 36% and 38% of the total deviance (D^2). First, the set of ecogeographical parameters available could not be correct. There could be other parameters that better explain the presence of the species. Second, some absences data determined in the field do not represent true absences. Species is still in expansion and probably many favourable habitats have not been colonized yet. Hence predictive model like GLM and GAM can not discriminate between true absence plot and those not yet colonised. The latter alternative seems not to be the case of GLM given that absences are more correctly predicted by the model than presences (Table 3.4).
But it has to be noted that explained deviance of GLM or GAM and the explained variance ENFA are not really comparable because they are calculated in different way. Indeed only the fact that ENFA is based on presence data decrease the probability of error. Moreover the explained deviance of GLM and GAM is not so surprising for an invasive species.
Habitat suitability map differences are not so big between GLM/GAM and ENFA, this means that considering only presence data, as ENFA does, do not give very different results.

If we compare the ROC values of the GLM with those from the GAM we see that GLM gives better predictions on a pseudo-independent data set, even if it explains a smaller part of the total deviance. GAM can explain a greater proportion of deviance because as a semi-parametric technique, it tends to better fit the data, but at the risk of a greater overfitting.
Finally for a management project GAM has preferably to be used, because explaining a little bit more of deviance than GLM and having a smaller uncertainty for highest presence probability.

4.3 Design based versus model based individual estimation

Thompson & Seber (1996) proposed a method to evaluate the number of individuals in a study area: it is supposed to be a current estimation of species density. In our results the design-based estimate is 5.259 mio of individuals. To have an idea of the importance of this value we proposed a comparison with an improbable species colonization of the whole study area, and we found that the current colonization estimate represent 3% of this value. The method of Thompson is directly based on the sampling, which it itself based on a model. This estimate takes strata in account. In our case the initial model (Benetollo & Monico, 2003), after field sampling, result to have little predictive value for our study area, because it was calibrated at the Swiss level and at a coarser resolution (Appendix D, Fig. 1). For these reasons, during the sampling we also found individuals in plot with probability of presence of only 0%-10%. Thus values obtained with Thompson's method risk to over-estimate the right value. However, the power of this analysis is to base the estimation on real presence data checked on field and to use them to weight the importance of strata.
Based on cumulating prediction values of number of individuals, we obtain much greater estimations for the model-based approach than the design based estimation. These estimations are supposed to be "at equilibrium" because they are based on the potential habitat suitability map. Hence a greater estimate is expected in this case. Furthermore there are some limitations to this type of estimation, which must be taken in account. For example, the availability of other ecogeographical predictors and of some other filter layers, (like landuse) and the fact that species will never reach some predicted habitat, for instance because of barriers to dispersal, can affect the estimation.

If we consider the mean number of individual per square observed in the field (corresponding at 49) we are able to calculate the number of pixel colonized by the species. Indeed with the design based estimation 107 345 pixels (12% of the entire area) are occupied. The same calculation is possible with the median value of individual per square observed in the field (30). In this case 175 330 (20% of the entire area) are colonized. The difference between mean and median is quite important, which means that the use of mean in this case is hazardous. These additionally results show the importance of species expansion.

4.4 General considerations

All results indicate that, in the PreAlps of Canton de Vaud, *Heracleum mantegazzianum* is still in its expansion phase. Pysek and Prach (1993) estimated the beginning of the exponential stage in central Europe to be in 1940. At the beginning species colonized mostly riparian habitat, but since 1970 it expanded massively into other habitat types, namely in settlements, along road, in verge and even in forests. Within the study area forests have been colonized only in the peripheral region, as indicated by distance to forest edge, and mostly where canopy is not too dense to prevent solar beams to reach the ground. Furthermore our estimations of population size suggest that the species could potentially double individual number in the study area in few years. And it is useless to say that this will have some repercussion on the environment and economy.

4.5 Preventions and controls

To prevent and control the spread of this noxious weed, some concrete actions should ideally be taken the next years. In areas which our models identified to be prone to host the invasive species, some preventive action could be underetaken.

First, it is important to manage the invasive plant proactively through continuous monitoring. To do this effectively, field staff should be trained in the identification of restricted and noxious invasive plants , collection of survey information, and the importance of destroying individual invasive plants and reporting new infestations in a timely manner.

Second, where possible, it could be usefull prevent unnecessary soil disturbances to limit the establishment of invasive plant infestations.

Furthermore practice due diligence by ensuring that all equipment, materials and vehicles are free of invasive plant seeds and plant parts before arriving on site. All agricultural implements or any equipment potentially exposed to invasive plants must be cleaned prior to use. Also equipment, materials and vehicles exposed to weeds are to be cleaned prior to leaving the infested site.

Finally communication between various stakeholder, Canton and Municipal Government agencies is beneficial to transfer information promoting regional awareness. Information such as the invasive plant history of certain locations or invasive plant infestation locations may be beneficial to all parties

For area already hosting the species current control methods are grazing, mechanical cutting and chemical treatment. Grazing method can be performed by cattle, sheep, pigs and goats (Caffrey

1994, Tiley & Philip 1994). In fields lightly or rotationally grazed by cattle, sheep and goats, mature plants may escape lethal damage and progress to flowering and fruiting stages. Pig foraging, however, would eradicate plants through damage to their roots (Tiley et al. 1996). Cutting of leaves is frequently used to reduce the above ground growth both in its vegetative and its flowering stages mostly on riverbanks and pathways. Cutting the vegetative plants above ground has only a cosmetic effect and does not lead to a long-term control. However if the tap-root is chopped with a spade or mattock below the base of the stem, this effectively kills the plant in both the vegetative or reproductive state. (Tiley & Philip 1992; Tiley et al. 1996). This treatment is suitable for small infestations or isolated plants, because labour-demanding. Pulling by hand, which requires glove protection, can eradicate occasional seedlings or very young plants (Dodd *et al.* 1994). Glyphosate, triclopyr, imazapyr are used in the United Kingdom as chemical control: when appropriately applied glyphosate can kill plants after 3 weeks (Williamson & Forbes 1982, Dodd *et al.* 1994), and imazapyr has a residual activity which prevents the germination of seedlings for several months. But it has been demonstrated that these substances provoke serious damages to the environment, namely on the photosynthesis of some algae (Anton *et al*., 1993).

In most cases successful eradication programmes need to be accompanied by prevention measures against re-colonization by the removed species and early warning systems should be put in place to detect colonisers early. A new infestation of the successfully eradicated species can be wiped out swiftly when detected early by using the appropriate eradication method, because the knowledge of the negative impact of the invasive species and the experience in controlling the species is established and will be supported by the stakeholders.

Monitoring of the impact of control actions needs to be put in place, preferably starting with small-scale activities to verify the impact of control operations, and if the results are not as expected, the management plan may need to be reconsidered and adapted in light of this new knowledge.

4.6 Perspectives

In further works adaptive sampling should be conducted based on a model fitted to the same area (not larger). If possible a higher number of squares should be sampled to improve models and have an independent data set for the evaluation (which is always better than an evaluation data set obtained by cross-validation (Guisan & Zimmermann, 2000)). Secondarily, new data could also give an idea of the degree of expansion. Moreover, the choice, if available, of other ecogeographical

parameters could improve models quality, and the use of more accurate filters, like landuse, could give better results in the model based estimation.

Furthermore it would be interesting to validate our models (GLM and GAM), to predict potential distribution of the species in other regions with the same characteristics, and check the predictions in the field (conducting eventually a new adaptive cluster sampling there).

5. Conclusion

The fact that *Heracleum mantegazzianum* is present in the Prealps of Canton de Vaud with high estimates for the number of individuals, and that it seems still in its exponential stage, suggests that many areas might still be prone to invasion. Not only biodiversity could be seriously affected but also public health. The attractiveness of the Giant Hogweed is a trap which could be avoided with efficient information campaigns. In the last years, a lot has been said on this matter, but until now few concrete actions have been undertaken. One aim of this work was to determine areas that are more prone to be invaded by the species and to estimate its total population size in the study area. Areas identified to host the species need absolutely some preventive measures to be taken in the next year and other suggested areas should further prospected. We believe very much that predictive maps can provide useful tools to support such prospection.

6. Acknowledgements

I would like to thank Antoine Guisan for accepting me in his team, for supervising the work and for his helpful suggestions. I'm particularly grateful to Simone Peverelli, Christoph Randin, Robin Engler and Thomas Czaka for availability and assistance with statistical analysis and informatics problem. A special thank go also to Nigel Yoccoz for interesting suggestions about the estimation of individual number.

7 Bibliography

Abbott, R.J. (1992). Plant invasions, interspecific hybidization and the evolution of new plant taxa. *Trends in Ecology and Evolution*. **7**: 401-405.

Anton, F.A., Ariz, M., and Alia, M. (1993). Ecotoxic effects of four herbicides (glyphosate, alachlor, chlortoluron and isoproturon) on the algae *Chlorella pyrenoidosa* Chick. *The Science of the Total Environment*, Supplement 1993, Proceedings of the Second European Conference on Ecotoxicology. Elsevier Science Publishers B.V., Amsterdam., pp. 845-851.

Baker, H. G. 1974. The evolution of weeds. *Annual Review of Ecology and Systematics* **5**: 1–24.

Benetollo, C. & Monico, N. (2003). Potentiel invasif de *Ambrosia artemisiifolia, Heracleum mantegazzianum* et *Senecio inaequidens* en Suisse. Travail de module de botanique. UNIL, Lausanne.

Blossey, B. &. Nötzold, R. (1995). Evolution of increased competitive ability in invasive nonindigenous plants: a hypothesis. *J. Ecol.* **83**: 887–889.

Brown, J.B. & Mitchell, R.J. (2001). Competition for pollinisation, effects of pollen of an invasive plant on seed set of a native congener. *Oecologia* **129**: 43.49.

Caffrey, J.M. (1994). Spread and management of *Heracelum mantegazzianum* (Giant Hogweed) along Irish river corridors. *Ecology and Management of Invasive Riverside Plants* (eds L.C.de Waal, L.E. Child, P.M. Wade J.H. Brock), pp.67-76. Wiley, Chichester.

Callaway, M. & Ridenour, W.M. (2004). Novel weapons: invasive success and the evolution of increased competitive ability. Front Ecol Environ; **2(8)**: 436 – 443.

Chittka, L. & Schürkens, S. (2001). Successfull invasion of a floral market. Nature **411**:653.

Clay, K. (2003). Conservation biology: Parasite lost. Nature **421**: 585-586.

Clegg, L.M. & Grace, J. (1974). The distribution of *Heracleum mantegazzianum* (Somm. & Levier) near Edinburgh. *Transactions of the Botanical Society of Edinburgh*, **42**: 223-229.

Copas, J. (1999). The effectiveness of risk scores: the logit rank plot. *Journal of the Royal Statistical Society Serie C – Applied Statistics*, **48**: 165-183

Dawe, N.K. & White, E.R. (1979). Giant Cow Parsnip (*Heracleum mantegazzianum*) on Vancouver Island, British Columbia. *Canadian Field Naturalist*, **93**: 83-83.

Davison, A.C. (1989). Model-Checking II: Binary data. *Revista Brasileira de Probabilidade e Estatistica*, **3**: 87-96.

Dodd, F.S., Waal, L.C. de, Wade P.M. & Tiley, G.E.D. (1994). Control and management of *Heracleum mantegazzianum* (Giant Hogweed). *Ecology and Management of Invasive Riverside Plants* (eds L.C.de Waal, L.E. Child, P.M. Wade J.H. Brock), pp. 111-126. Wiley, Chichester.

Efron, B. & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. *Chapman and Hall*.

Elton, C.S. (1958). The ecology of invasions by animals and plants. *Methuen.*

Ellstarnd N.C. & Schierenbeck K.A. (2000). Hybridisation and stimulus for the evolution of invasiveness in plants. *Proceedings of the National Academy of Sciences (USA).* **97**: 7043–7050.

Frontier, S. (1976). Étude de la décroissance des valeurs propres dans une analyse en composantes principales: comparison avec le modèle du bâton brisé. *J. Exp. Mar. Biol. Ecol.* **25**: 67-75.

Gigon, A. & Weber E.(2004). Report to the attention of the Swiss Agency for Environment, Bern.

Grime, J.P., Hodgson, J. & Hunt, R. (1988). Comparative plant ecology. A functional approach to common British species. *Unwin and Hyman*, London.

Grosholz, E. (2002). Ecological and evolutionary consequences of coastal invasions. *Trends in Ecology and Evolution* **17**: 22–27.

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

Hastie, T., Tibshirani R. (1986). Generalized Additive Models. *Statistical Science*. **1(3)**: 297-318.

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 (1998-2002). Biomapper2.1. Laboratory for Conservation, Biology, University of Lausanne.

Hoag, H. (2003). Invaders leave enemies behind. *Nature News Service*

Jeanmonod, D. (1999). La berce du Caucase: une genevoise belle, mais dangereuse. *Saussurea* **30**: 62-65.

Jeanmonod, D. (2002). Nouvelles stations pour Senecio inaequidens DC. ou la conquête irrésistible du séneçon du Cap. *Saussurea* **32**: 43-50.

Jehlik, V, & Lhotska, M. (1970). Contribution to the distribution and fruit biology of some synanthropic plants from the Pruho nice park, village of Pruhonica and Botic brook valley. *Studie CSAV*, Praha **1970/7**,45-95 (in Czech). in Pysek, P. & Prach, K. (1993). Plant Invasions and the Role of Riparian Habitats: A Comparison of Four Species Alien to Central Europe. Journal of Biogeography. **40(4)**: 413-420.

Joshi, C. M. Leeuw, J. D. & van Duren, I. C (2004). of Remote sensing and GIS applications for mapping and spatial modelling of invasive species. ISPRS, 12-23, Istanbul.

Lambrinos J.G. (2004). How interactions between ecology and evolution influence contemporary invasion dynamics. *Ecology*. **85(5)**: 2061-2070.

Lee, C. E. (2002). Evolutionary genetics of invasive species. *Trends in Ecology and Evolution* **17**: 386–391.

Lehmann, A., Overton, J.McC. & Leathwick, J. (2002). GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*. **157**: 189-208.

Lundstöm, H. (1984) Giant hogweed, *Heracleum mantegazzianum*, a threat to the Swedish countryside. *Weeds and weed control*. 25th Swedish Weed Conference, Uppsala, Vol.**1**, p.191-200.

Mandenova, I.P. (1950). Caucasian species of the genus *Heracleum*. Monographs, Ser. A, Systematics & Plant Geography. Tbilisi (in Russian).

McCullagh, P. & Nelder, J.A. (1989). Generalized Linear Models. *Second ed. London: Chapman and Hall.*

Mitchell, C.E. & Power, A.G. (2003). Release of invasive plants from fungal and viral pathogens. *Nature* **421**: 625-626.

Ochsmann, J. (1992). Riesen-Bärenklau (*Heracleum* spec.) in Deutschland: Morphologie, Ökologie, Verbreitung und systematische Einordnung. Diplomerbeit vorgelegt, Göttingen University.

Peterson, A.T., Papes, M. & Kluza, D.A. (2003). Predicting the potential invasive distribution of four alien plant species in North America. *Weed Science,* **51**: 863-868.

Pysek, P. (1991). *Heracleum mantegazzianum* in the Czech Republic – the dynamics of spreading from the historical perspective. *Folia Geobotanica et Phytotaxonomica,* **26**: 439-454

Pysek, P. & Prach, K. (1993). Plant Invasions and the Role of Riparian Habitats: A Comparison of Four Species Alien to Central Europe. Journal of Biogeography. **40(4)**: 413-420.

Pysek, P. (1994). Ecological aspects of invasion by *H. mantegazzianum* in the Czech Republic. *Ecology and Management of Invasive Riverside Plants* (eds L.C. de Waal, L.E. Child, P.M. Wade & J.H. Brock), pp. 45-54. Wiley Chichester.

Pysek, P., Kucera, T., Puntieri, J. & Mandak, B. (1995). Regeneration in *Heracleum mantegazzianum* – response to removal of vegetative and generative parts. *Preslia*, **67**: 161-171.

Pysek, P., Kopecky M, Jarosik V. & Kotkova P. (1998) The Role of Human Density and Climate in the Spread of Heracleum Mantegazzianum in the Central European Landscape. *Diversity and Distributions*. **4(1):** 9-16.

Reznick, D. N., and C. K. Ghalambor (2001). The population ecology of contemporary adaptations: what empirical studies reveal about conditions that promote adaptive evolution. *Genetica* **112/113**: 183–198.

Rhymer, J. M. & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*. **27**: 83-109.

Sokal, R.R. & Rohlf, F.J. (1981). Biometry: The principles and practice of statistics in the biological research. *W.H.Freeman & CO.*, New York.

Tiley, G.E.D. & Philip, B. (1992). Strategy for the control of Giant Hogweed (Heracleum manteggazianum Somm. & Lev.) on the River Ayr in Scotland. Vegetation Management in Forestry, Amenity and Conservation Areas. Aspects of Applied Biology. **29**: 463-466.

Tiley, G.E.D. & Philip, B. (1994). Heracleum mantegazzianum (Giant Hogweed) and its control in Scotland. *Ecology and Management of Invasive Riverside Plants* (eds L.C.de Waal, L.E. Child, P.M. Wade J.H. Brock), pp. 101-109. Wiley, Chichester.

Tiley, G.E.D., Dodd, F.S. & Wade, P.M. (1996). *Heracleum mantegazzianum* Sommier & Levier. *Journal of Ecology*, **84**: 297-319.

Thompson S.K., Seber G.A.F. (1996). Adaptive Sampling. *John Wiley & Sons*, INC.

Torchin, M.E., K.D. Lafferty, A.P. Dobson, V.J. McKenzie & A.M. Kuris. (2003). Introduced species and their missing parasites. *Nature* **421**: 628-630.

Vilà, M., Weber, E. & D'Antonio, C.M. (2000). Conservatiòn implications of invasion by plants hybridisation. *Biological Invasions*. **2**: 207-217.

Waal, L.C. de, Child, E.L., Wade, P.M. & Brock, J.H. (eds.) (1994). Ecology and management of invasive riverside plants. *John Wiley & Sons*, Chichester.

Williamson, J.A. & Forbes, J.C. (1982). Giant Hogweed (Heracleum mantegazzianum): its spread and control with glyphosate in amenity areas. *Weeds. Proc.* 1982 *British Crop Protection Conference*, 967-971.

Wyse Jackson, M. (1989). Observations on the Irish distribution of a plant with serious public health implications: Giant Hogweed (Heracleum mantegazzianum Sommier & Levier). *Bull. Irish Biogeogr. Soc*. **12**: 94-112.

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

Appendix A: Correlation table of quantitative predictors

Table 1: Correlation table of quantitative predictors

	mnt	precyy	sfroyy	slp	sradyy	swb	taveyy	topo	twi	riverdist	roaddist	builddist	foresteddist
mnt	1.00	0.97	0.60	0.62	-0.29	0.59	-1.00	0.54	-0.40	0.31	0.63	0.75	0.65
precyy	0.97	1.00	0.61	0.56	-0.32	0.63	-0.96	0.50	-0.37	0.25	0.62	0.72	0.65
sfroyy	0.60	0.61	1.00	0.45	-0.06	0.07	-0.58	0.48	-0.29	0.53	0.89	0.76	0.88
slp	0.62	0.56	0.45	1.00	-0.23	0.01	-0.61	0.62	-0.51	0.47	0.54	0.66	0.46
sradyy	-0.29	-0.32	-0.06	-0.23	1.00	-0.26	0.30	-0.15	0.08	-0.12	-0.10	-0.23	-0.09
swb	0.59	0.63	0.07	0.01	-0.26	1.00	-0.62	0.00	-0.03	-0.13	0.13	0.14	0.11
taveyy	-1.00	-0.96	-0.58	-0.61	0.30	-0.62	1.00	-0.51	0.40	-0.29	-0.63	-0.74	-0.64
topo	0.54	0.50	0.48	0.62	-0.15	0.00	-0.51	1.00	-0.55	0.54	0.51	0.56	0.58
twi	-0.40	-0.37	-0.29	-0.51	0.08	-0.03	0.40	-0.55	1.00	-0.23	-0.32	-0.43	-0.31
riverdist	0.31	0.25	0.53	0.47	-0.12	-0.13	-0.29	0.54	-0.23	1.00	0.57	0.44	0.60
roaddist	0.63	0.62	0.89	0.54	-0.10	0.13	-0.63	0.51	-0.32	0.57	1.00	0.77	0.80
builddist	0.75	0.72	0.76	0.66	-0.23	0.14	-0.74	0.56	-0.43	0.44	0.77	1.00	0.76
foresteddist	0.65	0.65	0.88	0.46	-0.09	0.11	-0.64	0.58	-0.31	0.60	0.80	0.76	1.00

Appendix B: ENFA validation

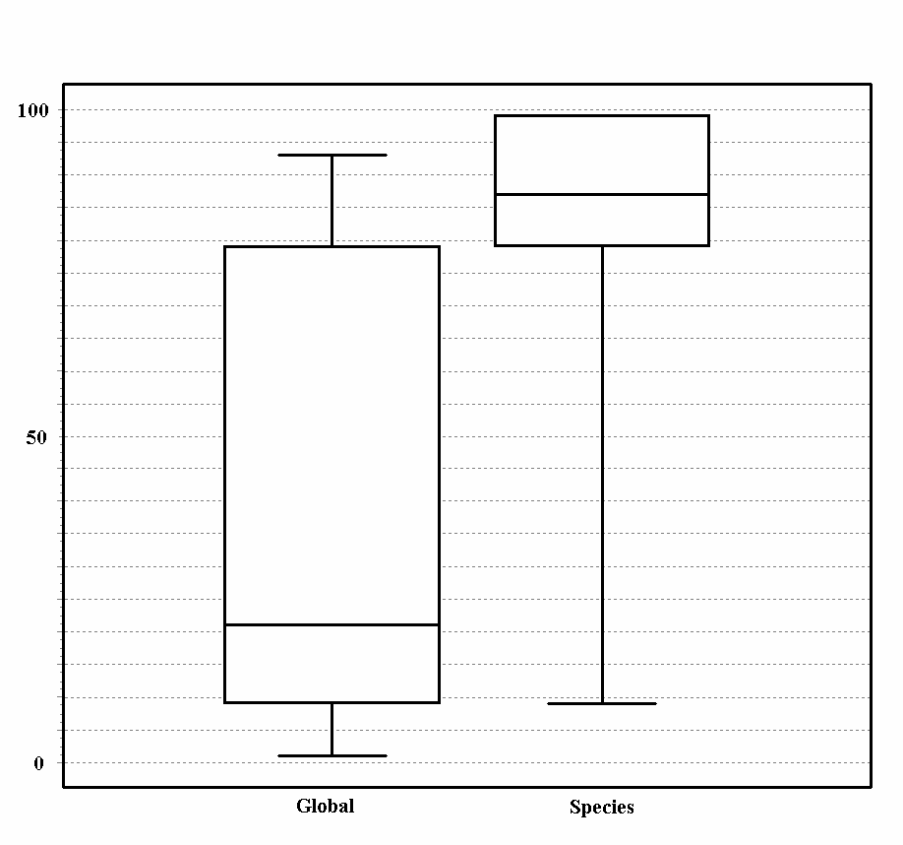

Fig. 1: Validation box-plot for the Ecological Niche Factor Analysis (Global Median = 21%, Species Median = 87%)

Appendix C: GLM model results

Table 1 : parameters retained in GLM, with the corresponding estimate, standard error, z value and z's associate probability

	Estimate	**Std.Error**	**Pr(>|z|)**	**Significance**
(Intercept)	-12.00	5.00	0.01632	*
foresteddist	2.44	$7.53*10^{-1}$	0.00118	**
$(\text{ foresteddist})^2$	$-4.22*10^{-1}$	$1.49*10^{-1}$	0.00472	**
slp	$2.86*10^{-1}$	$1.53*10^{-1}$	0.06237	.
$(\text{slp})^2$	$-8.49*10^{-3}$	$4.04*10^{-3}$	0.03561	*
taveyy	$3.16*10^{-2}$	$1.58*10^{-2}$	0.04481	*
$(\text{taveyy})^2$	$-2.72*10^{-5}$	$1.30*10^{-5}$	0.03609	*
topo	$1.02*10^{-2}$	$4.28*10^{-3}$	0.01738	*

Table 2 : ANOVA analyses

	Df	Deviance	Resid.Df	Resid.Dev	P(>|Chi|)
NULL			99	107.855	
foresteddist	1	7.3	98	100.555	**0.007**
$(\text{foresteddist})^2$	1	7.658	97	92.897	**0.006**
slp	1	2.131	96	90.767	0.144
$(\text{slp})^2$	1	8.383	95	82.384	**0.004**
taveyy	1	0.686	94	81.698	0.408
$(\text{taveyy})^2$	1	5.253	93	76.445	**0.022**
topo	1	6.9	92	69.545	**0.009**

Appendix D: Comparison of ENFA model calibrate at global and local level

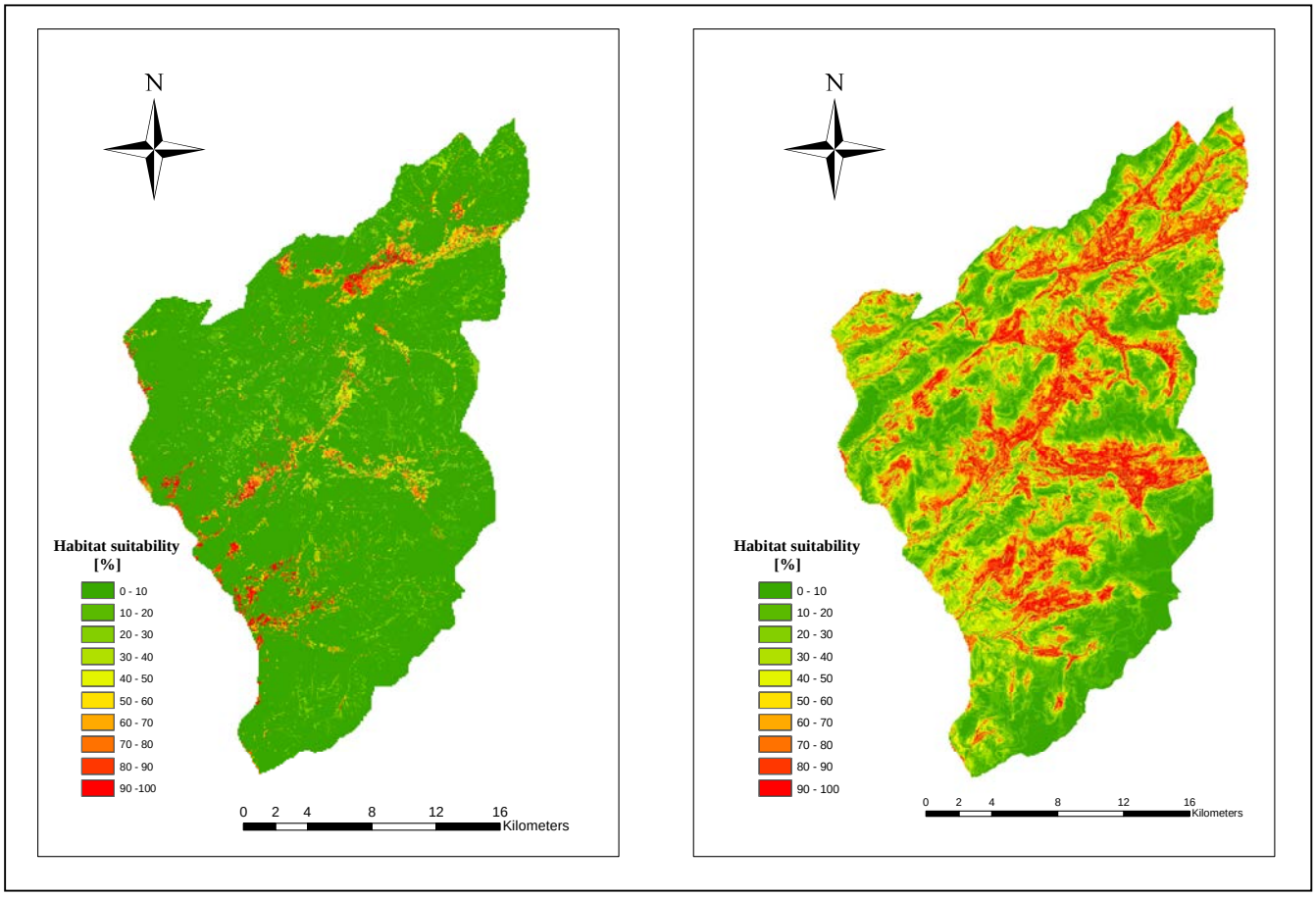

Fig. 1 : Comparison of ENFA model calibrate at global, Switzerland (left) and local level, Prealps of Canton Vaud (right)

Appendix E: Results of GLM and GAM models for the model based individual estimation

Table 1 : parameters retained in GLM, with the corresponding estimate, standard error, z value and z's associate probability

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	
(Intercept)	-23.4	11.90	-1.968	0.0491	*
slp	$7.99*10^{-1}$	$5.20*10^{-1}$	1.537	0.1242	
$(slp)^2$	$-2.63*10^{-2}$	$1.64*10^{-2}$	-1.606	0.1082	
topo	$7.26*10^{-3}$	$7.08*10^{-3}$	1.025	0.3052	
$(topo)^2$	$4.49*10^{-5}$	$5.67*10^{-5}$	0.792	0.4283	
$(foresteddist)^2$	$-1.06*10^{-4}$	$1.09*10^{-4}$	-0.979	0.3276	
precyy	$1.07*10^{-3}$	$6.48*10^{-4}$	1.658	0.0974	.

Table 2 : ANOVA analyses

	Df	**Deviance**	**Resid.Df**	**Resid.Dev.**	**P(>|Chi|)**
NULL			99	24.9582	
slp	1	0.7304	98	24.2278	0.3928
$(slp)^2$	1	3.6399	97	20.5879	0.0564
topo	1	0.2847	96	20.3033	0.5937
$(topo)^2$	1	0.0004	95	20.3029	0.9837
$(foresteddist)^2$	1	3.4891	94	16.8137	0.0618
precyy	1	4.0587	93	12.755	0.0439

For GLM the ROC approach gives an AUC' value of 0.51 and the bootstrap evaluation showed that the difference between estimated predictor values and the real population values is < $-1.38* 10^{-15}$

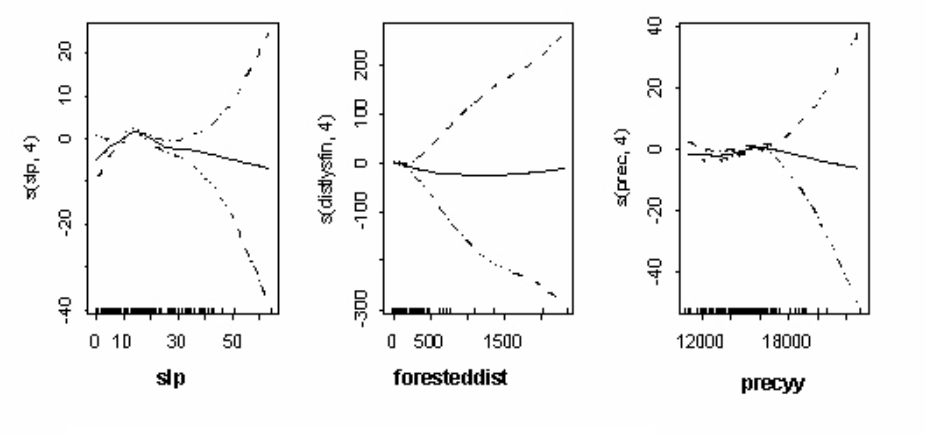

Fig. 1: GAM response curves for the three predictors

For non binomial functions GRASP use a simple correlation coefficient (r) as evaluation. In this case r = 0.85.

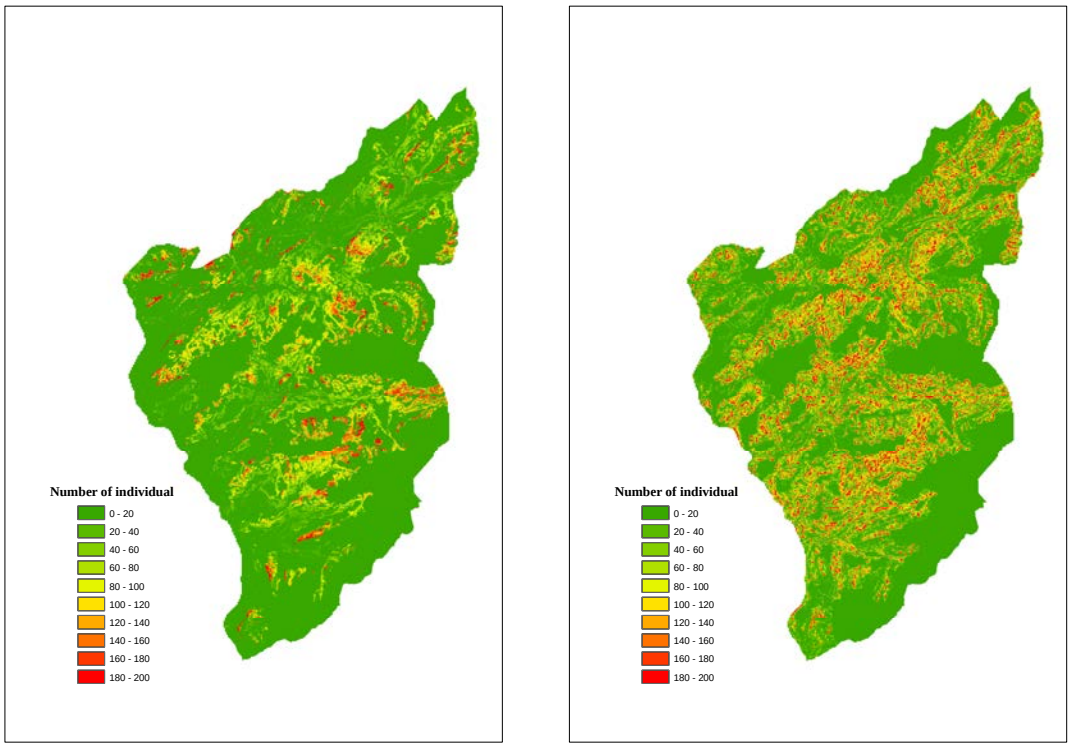

Fig. 2: Habitat suitability map of GLM (left) and GAM (right) models used to evaluate number of individuals

Appendix F: Black and Watch List of alien invasive species in Switzerland

Table.1: Black and Watch List of alien invasive species in Switzerland (Gigon & Weber 2004)

Black list	**Watch List**
Ailanthus altissima	*Amorpha fruticosa*
Ambrosia artemisiifolia	*Bunias orientalis*
Artemisia verlotiorum	*Cornus sericea*
Buddleja davidii	*Cyperus esculentus*
Elodea nuttalli	*Elodea canadensis*
Heracleum mantegazzianum	*Helianthus tuberosus s.l.*
Impatrients glandulifera	*Lonicera henryi*
Lonicera japonica	*Lupinus polyphyllus*
Ludwigia grandiflora	*Mahonia aquifolium s.l.*
Lysichiton americanus	*Prunus laurocerasus*
Polygonum ploystachyum	*Pueraria lobata*
Prunus serotina	*Sedum spurium*
Reynoutria japonica	*Senecio rupester*
Reynoutria sachalinensis	*Trachycarpus fortunei*
Rhus typhina	
Robinia pseudoacacia	
Rubus armeniacus	
Senecio inaequidens	
Solidago canadensis s.l.	
Solidago gigantea	