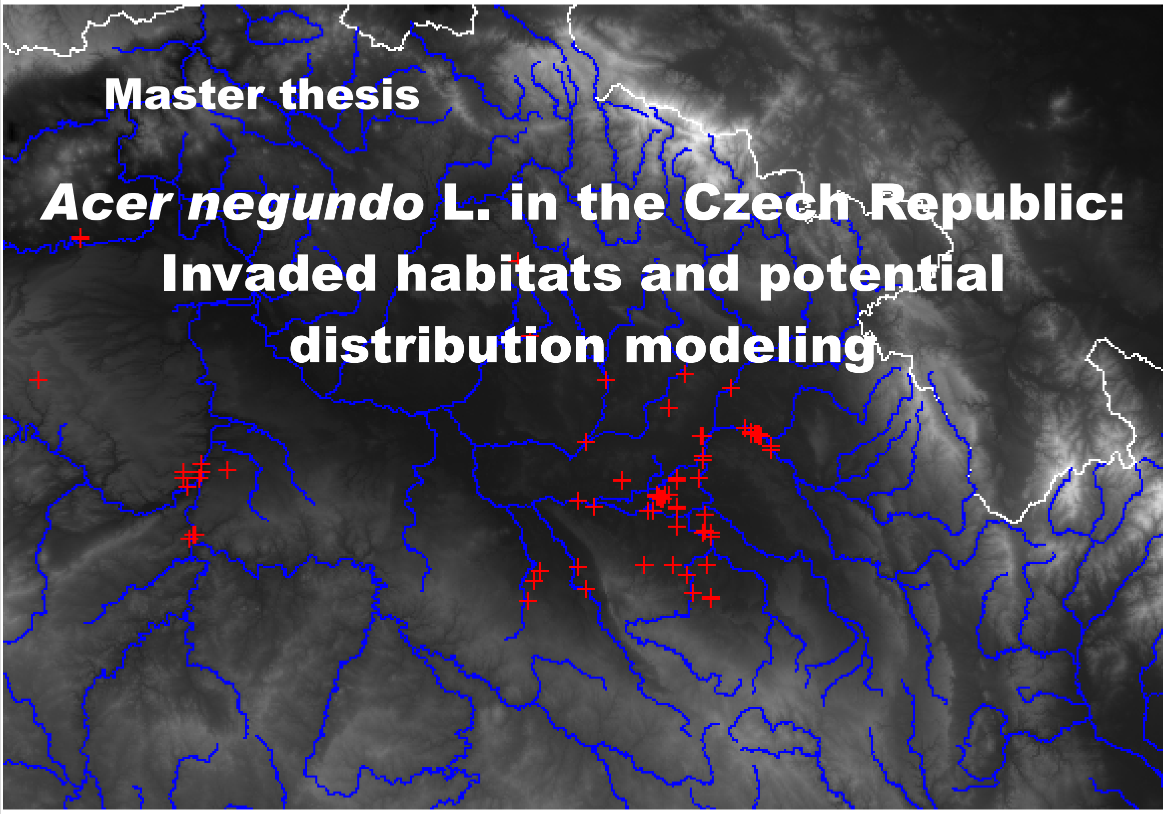
University of South Bohemia - Czech Republic

FACULTY OF BIOLOGICAL SCIENCES

Department of botany



Záboj Hrázský

Advisor: RNDr. Stanislav Mihulka, PhD.

České Budějovice 2005

Hrázský, Z. (2005): *Acer negundo* L. in the Czech Republic: Invaded habitats and potential distribution modeling. – Ms. [Masters Thesis, Department of Botany, Faculty of Biological Sciences, University of South Bohemia, České Budějovice, in English]

Annotation:

Habitats of *Acer negundo* wilding sites were described and documented by releves. Information about boxelder distribution was investigated using recently collected data from national mapping. Several models based on logistic regression (GLM, GAM) were fitted and prediction of potential wilding sites was performed. The predictors and types of model selection as well as prediction in different scales were discussed.

I declare I have elaborated this master thesis individually using only the cited literature.

....

Acknowledgments:

Data for this study were kindly provided by AOPK ČR and from The Czech Phytosociological Database.

I would like to thank to the following people for their help on different matters: Stanislav Mihulka for advising, David Zelený for valuable comments during the study, Petr Šmilauer for recommendation in the statistic field, Petra Konvalinková, Martina Filipová for comments on manuscript, Libuše B. Desetová for translation of geological terms and the most thanks due to Maggie Skodon for kindly improving my English.

This work would never be completed without the Biological Faculty of University of South Bohemia, without collective from the Department of Botany, without support from all people in DAPHNE CZ - Institute of applied ecology and without never ending patience of my wife.

Table of Contents

1. Introduction

1.1 Ecological invasions of plants

The process of species migration and exchange has taken place as long as species have existed. Colonization of new places and free niches and growth of newly established populations are natural and biosphere-wide. If such events are natural, what distinguishes the cases called ecological invasions?

The basic factor that characterizes the ecology of species invasion is its connection to human activities. Anthropogenic influence on vegetation has been occurring for at least 10 thousand years (since the neolithic period), when man as primitive agriculturalist (first as herdsmen and consequently as farmers) started to manage the landscape. At the beginning, the human impact was low and local, but the acceleration of cultural development led to unprecedented change in the dynamic of species exchange. Since the time when this acceleration has changed the rate of species exchange we have began to speak about invasions. There are several historical events in the last few centuries evidently connected with the increasing rate of ecological invasions. The most important of these are colonization of new continents after Great Discoveries (during the 15^{th} and 16^{th} cent.), expansion of urban areas and increasing rate of many types of disturbances (constructions, mines, 19^{th}-20^{th} cent.), and globalization of trade and personal movements in last 30 years (di Castri 1990). One of entry points for plant invaders can be botanical gardens. Europeans tended to carry their own plants to new colonized sites and bring home exotic ones in return. Some of the European neophytes were brought as decorative plants and later escaped into open country (*Reynoutria* spp., *Heracleum mantegazzianum*). Others were used as agriculture or technical plants (*Robinia pseudacacia*, *Helianthus tuberosus)*.

Secondly, a less evident but typical trait of ecological invasion is its impact on native species as well as expenses occurring when any expansion of invaders would need to be suppressed, e.g. for conservation or economic purposes. Non-native species increase the cost of managing cultivated landscape, where they can be detrimental weeds in agriculture, productive forestry or even in our gardens. They complicate nature conservation, when they occupy niches of native species. They can change habitat soil characteristics (Hadincová et al. 1997) and consequently decrease native species populations (Vitousek at al. 1997).

There are studies describing the progress of invasion in several phases (Kowarik 1995), and others using identification of these phases for estimating species- invasiveness. Unfortunately not all species formerly predicted as potential invaders act as invaders at all sites (Pyšek 2001). More than half of such cases stop at an earlier phase of invasion and do not need any further action. Such situations require more detailed studies of individual species, especially research in the field of automated prediction of potentially endangered sites.

1.2 Predicting invasions

The time-proved rule that prevention is often the most effective solution of problems with exponential curve of progress, is valid also in the case of inevitable biological processes like ecological invasions. This means any prospective works that allow the detection of invasions in their initial stages could be useful (Macdonald 1990, Richardson et al. 1989, or di Castri 1990).

A promising approach to such detection is to collect a database of invaded areas and establish a predictive system, which can identify the area of potential invasion . The next step is to choose the most effective method for monitoring such sites (Cronk and Fuler 1995). Thus if any initial phase of invasion is recognized an effective reaction can be undertaken. The first mentioned step - identification of the site - is often realized using predictive modeling. Modeling strategy and methods vary with the time and goals a given of study. Guisan and Zimmerman (2000) summarize various approaches and statistical methods, as well as cite many other less complex reviews, e.g. Franklin's (1995) review of exact statistical methods suitable for model-building or Brown's (1995) recapitulation of species abundance and distribution.

Linking the geographical information system (GIS) with a statistical model is the most useful tool in predicting species distribution (Hastings 1996) and can be also used in predicting endangered sites. The GIS is the input/output module of the system. Environmental variables and the presence/absence data in unified spatial scale represent input information for the statistical model. The probability of occurrence in the investigated site is achieved by comparing the combination of environmental variables in the occupied sites to those in all available sites. The rate of reproduction or dispersal strategies can be also included if there is any need for time-dependent response. Additional variables can be used the in case of an invasive process: level of disturbance, type of land-use, distance to potential travel corridors such as rivers, railways, or roads (Zalba et al. 2000). If all relevant information is covered by a model, the GIS layers with probability of occurrence changing in time can be released as a powerful tool in landscape management, nature conservation or research.

1.3 Predictive systems

There are several predictive systems which work as described above. This study uses GRASP - Generalized Regression Analysis and Spatial Prediction (Lehmann et al. 2002), an extension of statistical software based on GAM – generalized additive models (Hastie et Tibschirani 1986). The details of using GRASP are mentioned in the methodology section 3.5. There are also other systems which are often used: The Genetic Algorithm for Rule-set Production called GARP, which tries to find non-random correlations between the presences and absences of the species and the values of the environmental predictors, as well as form a set of rules, which is later used for prediction. Currently there are four types of rules implemented: atomic, logistic regression, climatic envelope and negated climatic envelope rules (Peterson 2001). BIOCLIM - A Bioclimatic Analysis and Prediction System - considers only climatic variables and is often used to prepare climatic datasets for other modeling systems (Busby 1986, Busby 1991). FloraMap - works on a continental scale and also considers only the climatic environmental variable (Jones et Gladkov 1999). MIGRATE, often used for plant dispersal modeling, is a cell based model using logistic regression (Collingham et al. 1996). Lastly there is Biomapper, which is based on Ecological Niche Factors Analysis (ENFA, Hirzel et al. 2002) - a mutation of PCA adapted for habitat suitability prediction (Hirzel et al. 2001). All these systems are summarized in Tab.1.

Tab.1: Overview of prediction systems and model types

System	Model	References	Example of use
GARP	set of rules: climatic envelope, regression, ANN	Peterson et Cohoon 1999 Peterson 2001	Stockwell et Peters 1999 Elith et Burgman 2001
GRASP	GAM	Lehmann et al. 2002	Zaniewski et al. 2002
BIOCLIM	linear regression, predictors envelope	Busby 1986 Busby 1991	Hutchinson 1989
MIGRATE	logistic regression	Collingham et al. 1996	Wadsworth et al. 2000
Biomapper	PCA, ENFA	Hirzel et al. 2001 Hirzel et al. 2002	Dettki et al. 2003 Patthey 2003
-	ANN - Artificial Neural Networks		Lek et al. 1996 Mastrorillo et al. 1997

Several systems based on multilevel predicting and self-learning such as GARP and the use of the artificial neural networks are able to give very precise prediction on one concrete dataset, but prediction rules are study-specific and further interpretation and generalizations are difficult (Guisan and Zimmerman 2000). Systems based on logistic regression or ENFA are suitable for precise prediction using the space of predictors with more variables. Only one model approach (GARSP) was performed in this study although a multi-approach comparison is more widely recommended.

1.4 Invasive trees in the Czech Republic

Eight trees in the Czech Republic are classified as invasive: *Acer negundo* L., *Ailanthus altissima* (Mill.) Swingle, *Fraxinus pennsylvanica* Marshall, *Pinus strobus* L., *Populus* × *canadensis* Moench, *Prunus serotina* Ehrh., *Robinia pseudacacia* L. and *Quercus rubra* L. (Pyšek 2002). Most of them were introduced as timber wood or anti-erosion vegetation cover. Common to all of them is fast growth and mass production of seeds. They are often planted in plantations and/or along roads so sources of seeds are widely dispersed. All of them, if occuring as mono-dominant, suppress the understorey herbs or change the habitat characteristics (free nutrients, soil reaction). Most of these trees are reported as also being invasive in Poland (Tokarska-Guzik 2003) and/or Germany (countries with comparable climatic region) as invasive too.

1.5 *Acer negundo*

Nomenclature

The Latin name *Acer negundo* L. is used in this study according to Kubát et al. 2002 (the usual common nomenclature codex in Czech republic since 2002). This name can be found in most papers dealing with the species as well as "boxelder" in English, though other names also exist: *Negundo aceroides* Moench. in Latin and others English synonyms such as: ash-leaved maple, black ash, cut-leaved maple, maple ash, negundo maple, Red River maple, Manitoba maple, stinking ash, sugar ash and three-leaved maple (Brink 1954).

Ecology

Acer negundo is a deciduous tree from the *Aceraceae* family, grows up to 25m height, and lives about sixty years. It is a fast-growing species and often occurs in multi-trunk, shrub-like form (Slavík ed. 1997). This growth habit corresponds to its natural biotope and life strategy: Typical habitats are moist soils in flooded areas, banks along rivers and lakes, or swamp margins. Boxelder is namely an understorey tree of floodplain forests. Successionally boxelder comes at initial phase following *Populus deltoides* and willows in new grounds in alluvial bottoms and persists to the middle stage in the understorey to be forced out by shading in later stages of succession. The forests where

Acer negundo naturally occurs can be found in eastern and central parts of North America. These alluvial forests are dominated by *Ulmus americana, Populus deltoides, Fraxinus nigra, Quercus palustris* , and *Salix sp*. (Dollar 1992).

Invasiveness

Boxelder is one of the 10% of neophytes in the Czech republic coming from the North America. It is one of 69 species classified as invasive in our country (Pyšek et al. 2003).

Boxelder has several life history traits commonly regarded as typical for invasive species. These traits are: mass production of well germinating seeds, effective dispersal mechanisms, short juvenile period and fast growth (compare Williamson et Fitter 1996, Noble 1989). Winged autorotatial seeds are useful for local dispersal in tens of meters. Vegetative spread from root sprouting enables long-term site occupancy and short-distance propagation (meters). River-flow or flood water intermediate long-distance transport of both propagules types: seeds and vegetative parts. Sprouting from 20 cm long and 2cm diameter boxelder sticks has been observed during field work and was reported by Komissarov (1964).

Occasional expansion in abandoned fields has been observed in boxelder‘s natural range. The adaptation to an initial phase of succession mentioned above makes boxelder a pioneer plant in many cases. (Maeglin et Ohmann 1973)

2. Aims

This thesis heads towards several goals consistent with the problematic described in the introduction chapter. First, the habitats where boxelder spreads spontaneously, are determinate and any co-occurrence species are identified. As suggested by boxelder's native habitat, riparian forests would most likely be the most commonly invaded type of vegetation. Evidence for this suggestion is collected from the available datasets.

Because all invaders are strongly related to landscape influenced by man, distances to roads, towns or railways are tested as possible predictors in boxelder distribution modeling. Furthermore, selected environmental variables are used in models of potential distribution. The types of information handling, the ways of deriving variables in GIS environment, as well as different scales are compared. These goals are extracted below.

1) To identify main habitats being invaded by boxelder
2) To evaluate any climatic or geographic variables useful for prediction of the invasion in geographical space
3) To compare a few approaches in using GIS
4) To evaluate built models among themselves and by another dataset

3. Methods

3.1 Datasets

Releves

The phytocenological releves were sampled in sites with sponaneous occurrence of *Acer negundo*. The extent of the releve, 10x10m, corresponds to the special character of several investigated sites: either ruderal stands along rivers and roads, or the abandoned land in industry areas. Some seminatural stands have been found in alluvial and rarely in other types of forests. All vascular plants were recorded in these plots. Their abundance was estimated in percentages with special categories (+ and r) for low density in the sense of Braun-Blanquet. A total of 171 releves were analyzed, including 70 new releves sampled by the author (2003, 2004) and 101 releves from The Czech National Phytosociological Database (Chytrý et Rafajová 2003).

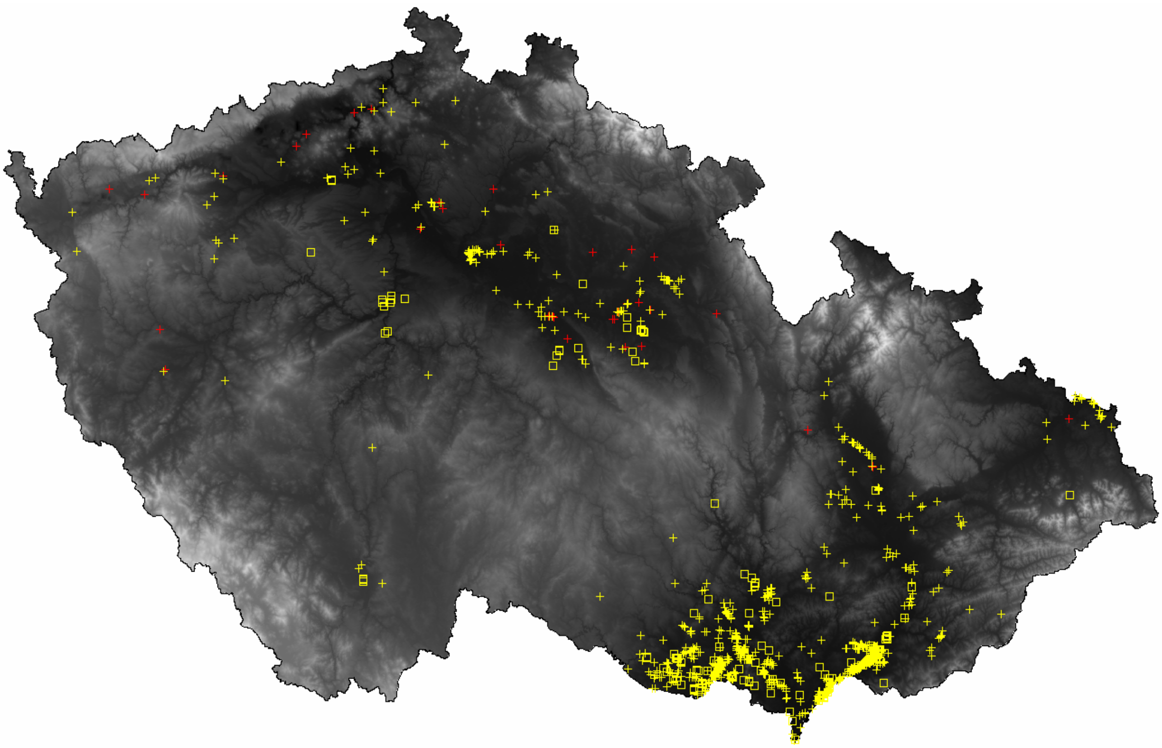


Fig.1. Distribution of *Acer negundo* in the Czech Republic. Yellow boxes are releves from the national database or collected by author, yellow crosses are records from national mapping where boxelder spontaneous spreads and the red crosses signals where is boxelder but spontaneous spread was not confirmed.

Presence data

The information on boxelder occurrences was obtained from three sources:

a) Most of the records come from recent **national vegetation mapping** (Guth 2002), which was performed by AOPK - Agency for Nature and Landscape Conservation in the years 2001-2004 and is the basis for European net of conservation areas NATURA 2000. This mapping covers all natural or seminatural landscape in the country and was performed for conservation purposes. Unfortunately the invasive species were not the target and this information was not collected by all of the experts. It means only presence data were available.b) The second part of presence-data comes from The Czech National Phytosociological Database (Chytrý et Rafajová 2003). Here both types of data are available, from natural and rural sites. Unfortunately this dataset does not cover all the country equally, see Fig. 2.

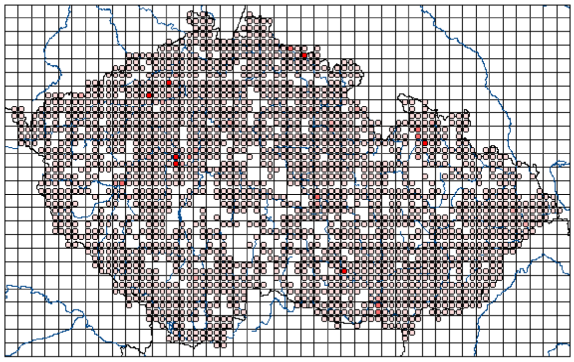


Fig.2. The distribution of releves in The Czech National Database. Empty area are those without any record in database.

c) The releves described above were sampled to cover all types of sites where *Acer negundo* has been spontaneously spreading. The sites were selected with special regard for anthropogenic history to complete dataset for all: natural, semi-natural, ruderal and urban sites.

An independent dataset was used for model evaluation. Presence/absence data were derived from the map (Slavík 1997, abbreviated as **slav** in graphs and tables) indicating squares of standardized mapping grid (see section 3.1 for details), where *Acer negundo* is present, whether planted or wilding individuals.

Pseudo-absence data

The absence of invasive species in its new areal can be caused by both, history of invasion process or unsuitability of the conditions at the site. So there are no true absence data available. The subset of pseudo-absence data was sampled randomly. Pseudo-absence were used for GARSP system, see modeling description in chapter 3.5, because this method requires the absence data for model fitting. The generated subset of sites was filtered through the presence data to get this pseudo-absence information. No random absences were allowed within a distance of 2,5 km around presence points - the distance was chosen arbitrary (compare Zaniewski et al. 2002). This absence-protected area partly deals with autocorrelation of species occurrence, which is common in ecology due the

Tab.2. Characteristic of variables used as environmental predictors. Scale types: R - ratio, C - circular, I - interval, O - ordinal, F - factorial, If the m is before the variable name, then it is an average for some area..

Variable	Abbr.	Scale	Unit	Description
altitude	alt	R	m	data from www.arcdata.cz rescaled to grid with100m square
slope	slo	C	°	calculated in GIS from altitude
southwestness	swn	I		calculated in GIS from altitude: cos(aspect-225°); in interval (-1;1) gives information about deflection from SW = 1
heat index	hix			calculated in GIS from altitude: cos(aspect-225°)*tg (slo)
mean annual temperature	mat	O	°	8 intervals represented by mean values: 2; 3,5; 4,5; 5,5; 6,5; 7,5; 8,5; 9,5
mean annual sum of precipitation	msp	O	mm	7 intervals represented by mean values: 450, 550, 650, 750, 900, 1100, 1300
distance to river	drv	R	km	calculated for grid with cell side 1km and 100m
distance to towns	dtw	R	km	calculated for grid with cell side 1km and 100m
distance to large towns	dci	R	km	calculated for grid with cell side 1km and 100m; only towns with more then 100 000 inhabitants
distance to railways	drw	R	km	calculated for grid with cell side 1km and 100m
distance to roads	dro	R	km	calculated for grid with cell side 1km and 100m
distance to national road	drn	R	km	calculated for grid with cell side 1km and 100m
potential vegetation	mpv	F		Neuhäuselová 1998, 44 levels, see Tab.7 in appendix
geology	geo	F		20 levels, see Tab.8
land cover	lcv	F		4 levels of land cover (buildings, forest, agricultural land, water)

factors like reproduction and dispersion (Guisan and Zimmerman 2000).

Environmental data

The information about environment was collected from several sources and added to a GIS database. Tab.2 describes the basic characteristic of the environmental dataset and Tab.3 shows characteristics of 3 grids in different scales, which were used in the study: grids with squares 100m, 1km and cca 12x11 km. The last one is the standardized grid for Middle-European flora and fauna mapping (in text as **standardized mapping grid**) and is defined as 10x6 geographical minutes.

Basic topographic information was derived from the digital elevation model (DEM) with 100 m steps. The elevation, the slope and the „southwestness“- number in the subrange of (-1;1), which gives the information about deflection of the slope from the warmest orientation and the combination of slope and southwestness in heat index: hi = cos(southwestness)*tg(slope).

The effects of three factorial variables were tested: potential vegetation types (Neuhäuselová 1998) - 52 categories see Tab.7 in appendix for description, 20 classes of geological substrate (map in 1:500 000 scale, see Tab.8 in appendix) and 4 land cover classes including urban areas, agricultural landscape, forests and water.

Furthermore the relations between *Acer negundo*-wilding sites and several potential anthropogenic or natural vectors - rivers, roads, railways and the spatial pattern of cities - were investigated. The distances to these features were calculated in two ways: in vector layers as distance between points, lines and boundaries and in a grid with 100m squares. The first approach allows to get distances with high accuracy for exploring relations in real local scale. The second one is useful for landscape scale prediction of boxelder occurrence.

Tab.3. Spatial characteristics for three different scales used in the study

square site	N. of cells total	N. of cells presences
100 m	7876200	950
1000 m	78862	592
12 x 11km (standardized mapping grid)	678	144

3.2 GIS analysis approach

Principally two approaches were used for spatial analysis - the point one and the polygon one:

a) In the point approach the information was taken from thematic layers only where points of presence or absence were situated. It means the variable for one point gets the same value as the cell of the grid layer in which that point falls. The matrix of 11 continuous variables and 3 factors was built in this way.

b) The polygon method accesses the area around the target points (circles of 6555 m radius) or an area in a cell of a large grid such as squares ot the standardized mapping grid. The ratio scale variables were calculated as means of cells falling in the target area. Factor variables were divided into dummy variables, where each new variable got the value of an area fraction occupied by one level of the factor. The sum of the new dummy variables for each factor (geo, mpv or lcv) is 1. (e.g. from the sample at Fig.3 just three of 20 dummy variables which originate in the geological layer have taken a value other than zero). The matrix of 11 calculated variables (columns of means) and 68 factor replacing ratio-scale variables was obtained.

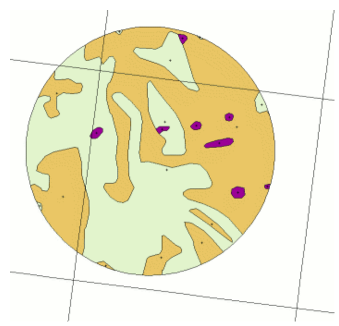


Fig.3. Circle sample of the geological layer around the point of boxelder occurrence. A standardized mapping grid cell is also displayed.

Datasets for prediction were accessed in two scale levels. In the 1000m grid the prediction was calculated directly from thematic layers. Values for squares in the standardized mapping grid were derived by the polygon approach described above.

The effects of individual predictors were tested in the most accurate scale - a 100m grid. Unfortunately these high-precision data are too large for handling with commonly used statistical tools (more than 100 000 cases) and so prediction was not performed in this scale.

3.3 Releves analysis methods

The set of releves was classified by commonly used TWINSPAN software (Hill et al. 1975) with default options in three dividing levels. More precise classification was not performed, because high heterogeneity between releves caused unequal divisions with one or two releves in one class. This was not useful for further analysis. Obtained classes were used for easier interpretation of diagrams from ordination methods. Canonical analysis (CA ter Braak 1986) was used to describe the species composition of vegetation at boxelder wilding sites. Analysis was made in The Canoco for Windows and CanoDraw software (ter Braak et Šmilauer 2002).

3.4 Exploratory analysis of predictors

The effects of selected individual predictors were tested by t-test comparing means of values in two groups - presences sites and pseudoabsences sites. Because of geographical relations in data the individual samples were not completly independent and therefore a non-random permutation test was performed - 10,000 permutations per set (Urban 2003). The distribution of values at presence and absence sites is showed in Q-Q plots. Dataset for this initial exploration was gained from a 100 m grid (1900 selected cells: 950 presences and 950 absences). All these graphs and tests were realized through the R v. 1.9.1 software.

3.5 Modeling methods

Before any models were built the correlations between predictors were assessed through the correlation matrix - see appendix, Fig.17.

All predictions in this study were performed in statistical software R (v. 1.9.1) using GRASP package, a set of R-language functions developed to facilitate the modeling and analysis of species spatial distributions (Lehmann et al. 2002). Distribution was modeled by GLM and GAM, using a logistic link and a binomial error term. All models were fitted with the predictor variables listed in Tab.2 using a both-directional stepwise procedure to include only those variables whose contribution to explained deviance was significant - tested by χ^2 with minimal p-level <0.01. Variable contribution was evaluated by assessing the variation in residual deviance as predictor variables were sequentially added and then dropped from the model. Models were selected by following strategies:

a) with maximum predictors, which could significantly improve the explained deviance, these are marked as **gam_b** and **glm_b**.

b) “the minimalistic“ versions of the previous two models were built omitting the variables, which contributed to an explained deviance of less than 2%. This is useful when interpreting models. These models were called **gam_m** and **glm_m**.

c) the fifth model was glm with only one predictor alt - **glm_a**, because altitude had contributed about 50% of the explained deviance and there was an assumption only this predictor is able to describe the distribution of boxelder wilding sites.

These foregoing models were all fitted with data from 1km grid. Another model was performed using the data derived from standardized mapping grid for comparison between different scales. This last one was named **B_gam** and is equivalent of gam_b**,** because the best explained deviance rule was used for model building.

The six models mentioned above and **slav** dataset were compared by **ROC** statistic and in graph plotting the numbers of true predicted values versus false predicted values.

The **ROC** curve analysis is a wide-ranging subject, with many different methods for estimating and comparing curves. The implementation in this study uses the non-parametric method for constructing curves as described in (Beck et Schulz 1986). The Hanley and McNeil method (1983) was used for comparing curves, but if the models predictions are highly correlated (Pearson's correlation coefficient $r > 0.9$) or the area under the curve is outside 0.7 to 0.975 the 2 curves can not be compared. The Analyze- it, v. 1.71 software was used for the above statistics.

4. Results

4.1 TWINSPAN and ordination analysis of releves

Classification analysis led in five unequal groups of releves: marshes, alluvial forests, moist meadows, dry meadow, secondary pine forests (Fig.4). These groups were named according to generalized habitat characteristics indicated by species composition. The pattern of the releves is better shown in Fig.5, where also division-causing species are listed.

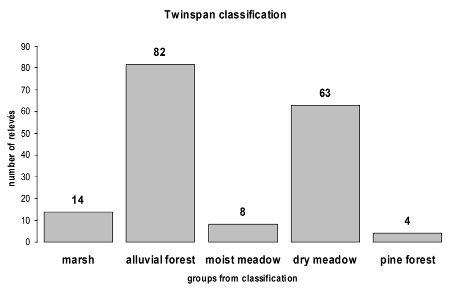


Fig.4. Number of releves in the TWINSPAN classification groups named by stands they probably occupy. The moisture gradient is shown from left to right.

The **marsh** group contains small area plots from ecotone of the ponds at two localities. It is usual that this type of vegetation occurs in small patches around water areas in the Czech Republic and therefore only a few releves are taken from these sites.

The **alluvial forests** group contains a wide range of alluvial woody vegetation from seminatural bird cherry-ash woodland (*Pruno-Fraxinetum*) through poplar-pedunculate oak woodland (*Querco-Populetum*) and white willow woodland (*Salicion albae*) to the marginal, but intensively invaded type of alluvial woodland in the Czech republic the Pannonian elm-ash woodland (*Fraxino panoniceae - Ulmetum*). The secondary poplar woodlands with dominant *Populus canadensis* are also included in this group.

The meadows are divided into two specific groups: As is natural for invasive species , plenty of wilding areas are sites in the different succesional stages from field or meadow to forest. Therefore “meadow“ in this study is used in the sense of an abandoned meadow, where succesion takes place with *Acer negundo* contribution. A high number of releve have fallen into the group of **dry meadows** which are derived from abandoned fields or mesophile meadows from the *Arrhenatherion* union. Such stands are dominated by *Elymus repens*, *Calamagrostis epigejos, Arrhenatherum elatius,* and or *Dactylis glomerata;* and the presence of scrubs, seedlings and juveniles of woody species indicates the successional change. The second group of meadows are **moist meadows** derived from alluvial meadows (*Alopecurion* union).

The last one is the pine forests group, which involves secondary forests at dry stands recently dominated by *Pinus sylvestris*.

Fig.5 gives the best understanding of how releves are distributed among the vegetation types mentioned above. Releves in the diagram are unequally distributed along three environmental gradients. The most evident one is the moisture gradient from the dry stands on the left to the marshes on the right. Second gradient is the continuum of "forest to meadow" species change, which could be interpreted as gradient of successional stages. The species composition also indicate that both ruderal and seminatural sites are invaded by *Acer negundo.*

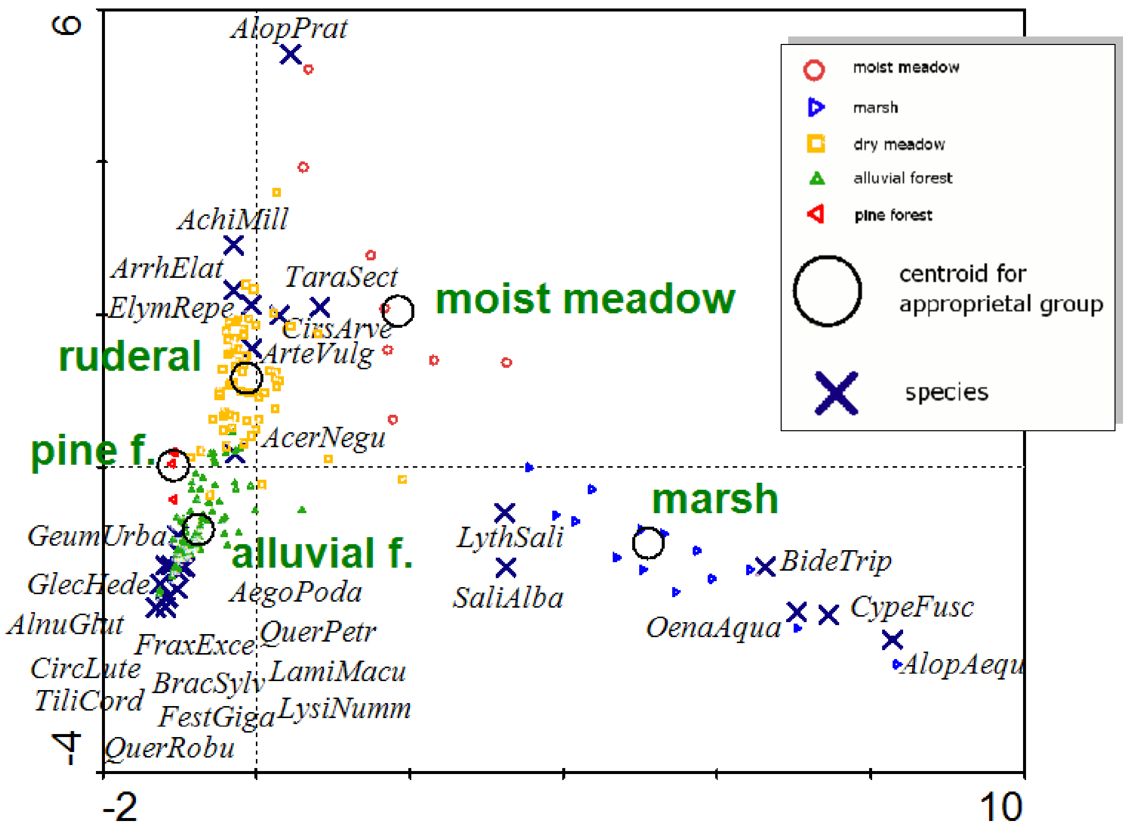


Fig.5. Ordination diagram from CA. Small signs mark releves, crosses mark species and larger circles are centroids for groups obtained by classification. Membership of releves in groups is indicated by different colors and shapes of signs.

4.2 Habitats in national vegetation mapping

The habitat classification from *Habitat Catalogue of the Czech Republic* (Chytrý et al. 2001) was used in national vegetation mapping. Habitats are determined by the diagnostic and dominant species as well as by a stand‘s characteristics. Such classification corresponds to the classical phytosociological units at union or association level, or several associations are grouped into one widely defined unit.

The most often invaded habitats are L2.3 and L2.4 - both are alluvial forests mid- or down-stream of great rivers (Fig.6). L2.3 signals hardwood forests of lowland rivers (*Ulmenion* Oberdorfer 1953:) and L2.4 are willow-poplar forests of lowland rivers (*Salicion albae Soó 1930)*. Descriptions of others habitats can be found in Catalog of habitats (Chytrý et al. 2001). Note that all four associations of oak-hornbeam woodlands in the Czech republic are affected by invasion of *Acer negundo* (L3.1, L3.2, L3.3, L3.4).

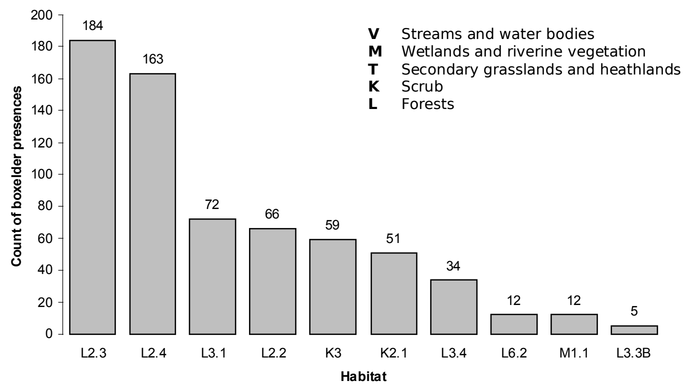


Fig.6. Habitats the most often invaded by Acer negundo. Habitats with less then 5 occurrences: K1, K2.2, L1, L3.2, L5.1, L5.4, L6.4, L7.1, L7.2, L7.4, M1.4, M1.5, M1.7, T1.1, T1.4, T1.5, T3.5B, V1G.

There are only seminatural forests described above, while the half of releves from chapter 2.1 come from non-forest or secondary forest sites. These sites were in national vegetation mapping marked as “X“ habitats. There were found more than 50 cases of boxelder wilding in four of them : **X12** stands of early successional woody species, **X7** herbaceous ruderal vegetation outside human settlements, **X8** scrub with ruderal or alien species, and **X9B** forest plantations of decidous allochtonous trees.

Acer negundo was found in 998 polygons in the whole Czech republic. Almost two hundred of them were omitted from the potential distribution analysis because it was impossible to distinguish between trees of wild or cultural origin.

4.3 Exploratory analysis of predictors

Eleven continual variables have been proved as predictors of *Acer negundo* wilding sites. Basic characteristics are listed in Tab.4:

Tab.4. The basic characteristics of the predictors for sites, where wilding *Acer negundo* was observed. Names of the predictors are explained in Tab.2.

	alt	slo	swn	mat	msp	drv	dtw	dci	dro	drn	drw
Min.:	125	0.010	-1	5.5	450	0	0	0	0	0	0
1st Qu.:	173	0.605	-0.71	8.5	450	0.1	1.2	15.2	0.4	1.1	0.8
Median:	187	1.205	0.14	8.5	550	0.7	2.35	30.8	1	2.5	1.6
Mean:	205.7	1.760	0.04	8.83	523.4	1.71	3.07	27.03	1.2	3.58	2.17
3rd Qu.:	222	2.161	0.75	9.5	550	2.68	4.3	37	1.7	5	3
Max.:	592	15.778	1	9.5	1100	13.8	13.7	70.7	8.5	18.5	8.8

Arithmetic averages of the subsets for presences and absences of individual variables were compared by t-test and results were evaluated by permutation test. The averages of presence/absence data subsets were different for almost all of used variables (Tab.5) and most of the achieved t-test statistics were significantly (p>0,001) different from t-statistic acquired in t-test performed on the same dataset with randomly permutated cases. The only exception was **swn**, when the difference between averages was significant only at p>0,05 level (Fig.7 compares high significant **mat** and low significant **swn** as an example). Distributions of values - in particular presence/absence dataset - are displayed in Fig.8. Mean annual temperature (mat) and mean annual sum of precipitation (msp) show the clearest difference in the distribution of values between presence and absence data subsets. Altitude and slope seem both to have a very strong

Tab.5. Summary of averages between the presence/absence data subsets and t-test statistic.

variable	µ - arithmetic mean absences	µ - arithmetic mean presences	t statistic	p-value
alt	206.000	429.000	-41.500	0.000
slo	1.760	4.200	-19.060	0.000
swn	0.043	-0.030	2.340	**0.019**
mat	8.835	7.485	38.620	0.000
msp	523.421	657.366	-28.580	0.000
drv	1.716	3.090	-13.020	0.000
dtw	3.066	4.471	-10.980	0.000
dci	27.028	30.844	-5.260	0.000
dro	1.195	1.426	-4.960	0.000
drn	3.581	4.521	-5.690	0.000
drw	2.168	3.764	-13.960	0.000

negative effect on boxelder occurrence. The group of “distance-from“ variables exhibits trend of increasing difference with increasing distance from the appropriate landscape feature. Note that distance from a river (drv) changes the rate of this increasing trend at the beginning. Southwestness (deviance from ideal sun

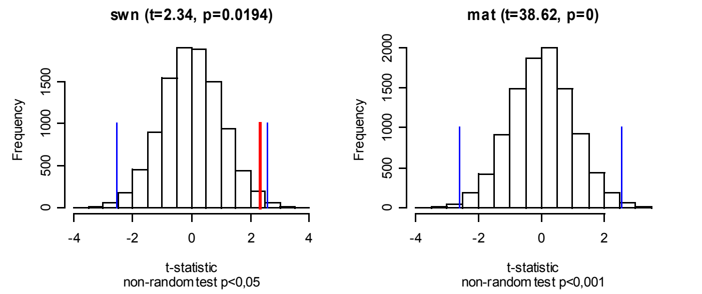


Fig.7. Distribution of t-test statistic after 10 000 permutations of cases, for two variables only from 11 tested: high significant mat and low significant swn. The down written p-value is achieved significance by comparing randomized_permutation_t-test_values distribution to the t-test statistic for non-randomized data, which is explicit above the histograms and shown by red thick line. The blue lines are the 0,5% and 99,5% quantiles.

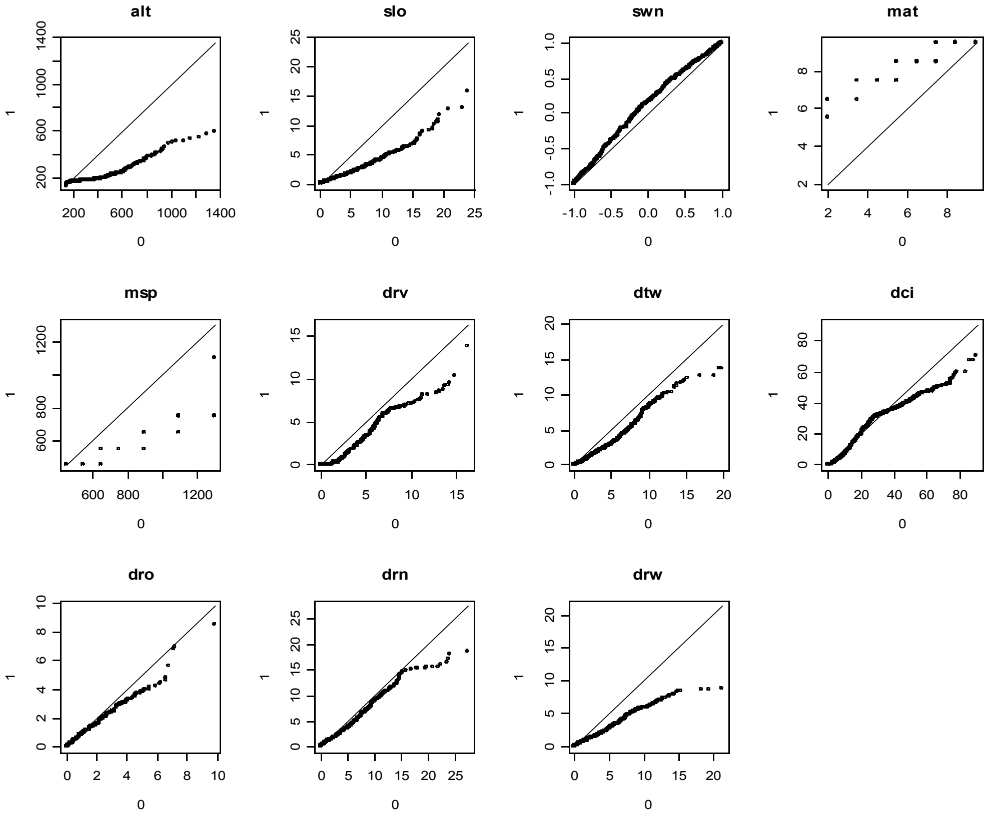


Fig.8. Q-Q diagrams of environmental variables. Two subsets of data (from presence sites and from pseudo absence sites) were sorted ascending and plotted against themselves. Deviation from the diagonal line signals the difference in distribution of values in data subsets. The isolated points in the graph for **mat** and **msp** are effects of interval scale for these variables (see Tab.2).

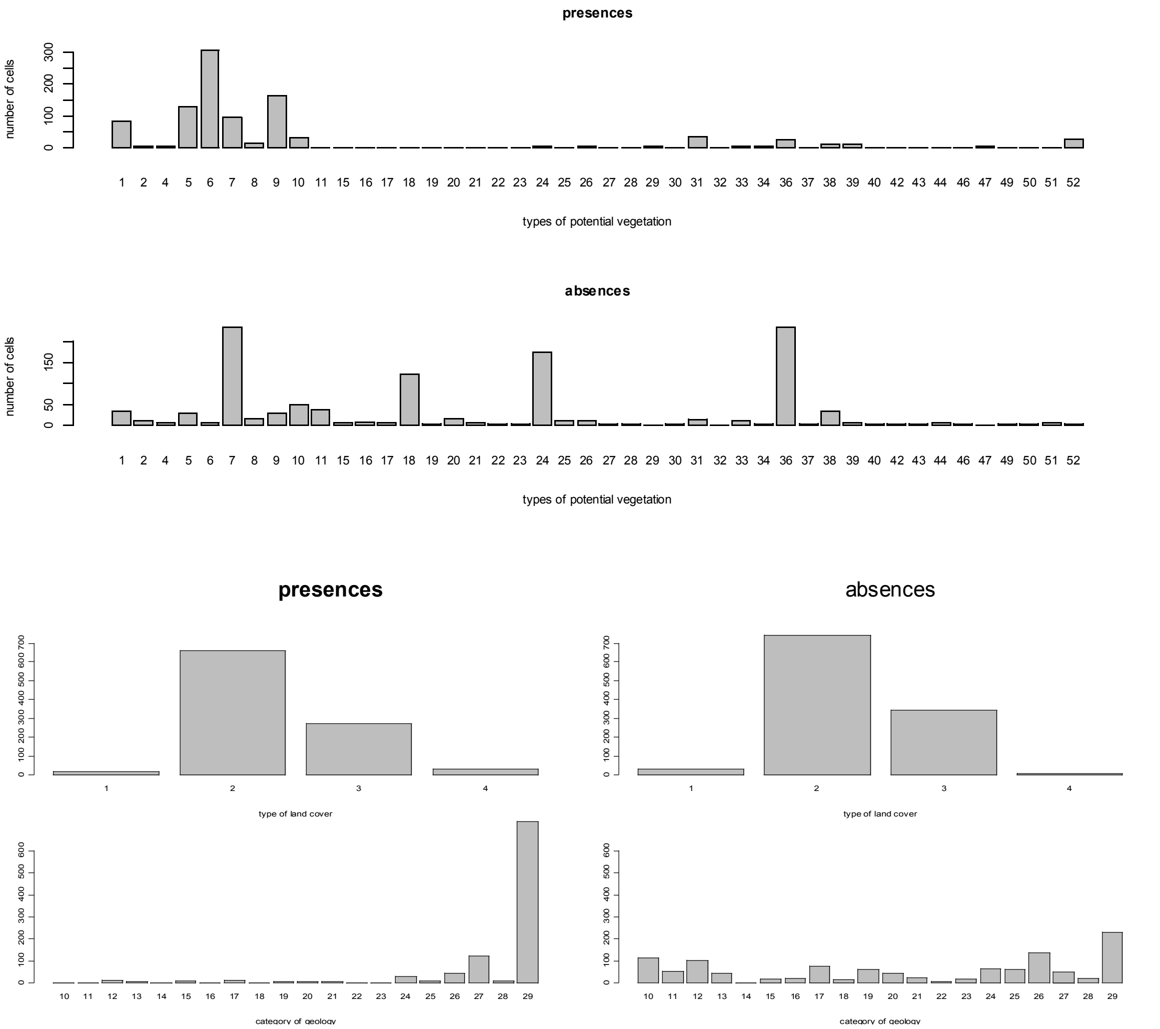


Fig.9. Participation of individual factor levels for four used factor variables. Situation for sites with Acer negundo occurrence is presented in left column and for sites without it in right column. The upper are categories of potential vegetation, middle land cover classes, and down are categories of geology. For legend on geology or vegetation units see appendix Tab.7 and Tab.8. The land cover classes are 1-urban area, 2- agriculture landscape, 3- forests, 4- water. Total number of cells was 950 for each presence or absence subset.

exposure) deviates from both-side equivalence only between upper and down quartiles and the extremes do not differ.

The factor variables were examined through the proportion of their levels in both presence and absence subsets of data. First the bar plots (Fig.9, page 21) were drawn individually for each subset and in the case of derivate dummy variables were compared to the response versus predictor plots (appendix Fig.15). From both approaches the same pattern is evident for potential vegetation units. There are two main points in these plots: alluvial forests (1, 5, 6) evidently increased their proportion in the presence-subset, conversely common abundant vegetation units such as oak-hornbeam woodland (*Melampyro nemorosi-Carpinetum*), acidophilous oak or beech woodlands (*Luzulo albideae-Quercetum, Luzulo-Fagetum*) as well as rich beech woodlands (*Dentario eneaphylii-Fagetum)* decrease their proportion in the presence subset (7,18,24,36). Less evident is difference in land cover, where an increasing trend can be found in bar-plot for the water category and decreasing in the urban category. This difference is only marginal, when the proportion of land cover types in the circle samples around points is investigated. The complementary exchange of dominant landscape features (fields versus forests) can be seen much better in response versus predictors plot (appendix Fig.16). The geology categories exhibit the same trend as potential vegetation, but are unequally divided into two groups:

a) with increasing proportion (27 and 29 - sediments as quaternary loam, loess, sands, gravel/broken stone and tertiary sands and clays) at *Acer negundo* wilding sites and
b) with the decreasing proportion including all other bedrock types.

Before models fitting the collinearity in datasets was investigated through the correlation matrix (Fig.17 in appendix). The strong correlation to altitude was observed in climatic variables (temperature, precipitation) as well as in “distance-from” variables and factors.

4.4 GAM and GLM models

Six models were fitted as described in section 3.5. They differ in family (GLM,GAM) and in complexity. All the models show between 49-61 percent of explained deviance. Individual predictors and the number of degrees of freedom for a spline smoother if used - in the case of GAM - are displayed in Tab.6. The most often used predictors were mpv (potential vegetation), mat (annual mean temperature) and alt (altitude). Because of the strong correlation among predictors driven by altitude, most of them were excluded from models in terms of non increasing proportion of explained deviance.

Tab.6. Summary of fitted models. Models compared in this table are: prediction in the scale of standardized mapping grid (**B_gam**), GAM and GLM models in a grid 1km x 1km (**ROC$_{1000}$**) and their averiges for squares in the standardized mapping grid (**ROC$_{bot}$**), suffix **_b** means best fit model, **_m** indicates minimum predictors model, and **_a** means model with single predictor alt (for details see section 3.5), in addition the map from Slavík 1997(**slav**) is compared by ROC. **D²** represents proportion of deviance explained by the model and used predictors are specified with formula indicating when the spline smoother **s()** was applied and number of degree of freedom for spline smoother. Models are sorted by increasing ROC$_{bot.}$

	Bpred	**glm_m**	**gam_b**	**glm_a**	**gam_m**	**glm_b**	**slav**
D^2 [%]	49,4	59,2	60	49,1	57,4	61,1	
Predictors	s(mmat, 4)	mpv	mpv	alt	mpv	mpv	
	s(mmsp, 4)	mat	s(alt, 3)		s(alt, 3)	alt	
	s(mdtw, 4)	geo	s(mat,3)			mat	
	s(mdci, 4)		drw			drw	
	s(mdro, 5)		s(dro, 2)			geo	
	s(v7, 4)						
	s(v24, 4)						
ROC$_{bot}$	0,931	0,920	0,916	0,913	0,907	0,889	0,645
Bpred		0,2274	0,1001	0,0515	**0,0219**	**0,0002**	**<0.0001**
glm_m	0,2274		NA	NA	NA	**0,0005**	**<0.0001**
gam_b	0,1001	NA		NA	NA	**0,0023**	**<0.0001**
glm_a	0,0515	NA	NA		NA	**0,0139**	**<0.0001**
gam_m	**0,0219**	NA	NA	NA		NA	**<0.0001**
glm_b	**0,0002**	**0,0005**	**0,0023**	**0,0139**	NA		**<0.0001**
slav	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	
ROC$_{1000}$		0,950	0,952	0,926	0,945	0,954	
glm_b		NA	NA	**<0.0001**	NA		
gam_b		NA		**<0.0001**	NA	NA	
glm_m			NA	**<0.0001**	0,2217	NA	
gam_m		0,2217	NA	**<0.0001**		NA	
glm_a		**<0.0001**	**<0.0001**		**<0.0001**	**<0.0001**	

The success of prediction was evaluated by ROC statistic in two scales:

a) in a standardized mapping grid for all six models and the dataset from another source

b) in 1km grid for five of them excluding Bgam model

ROC curves are shown in Fig.10. They are made by introducing a cut-off somewhere in the range of the output (0;1) - to classify continual probability into presences and absences - and compared these to the true situation given by the dataset of real presences and absences. Each cut-off corresponds to a point on a ROC curve. The ROC curve has the sensitivity (the probability of successfully predicted presences) plotted

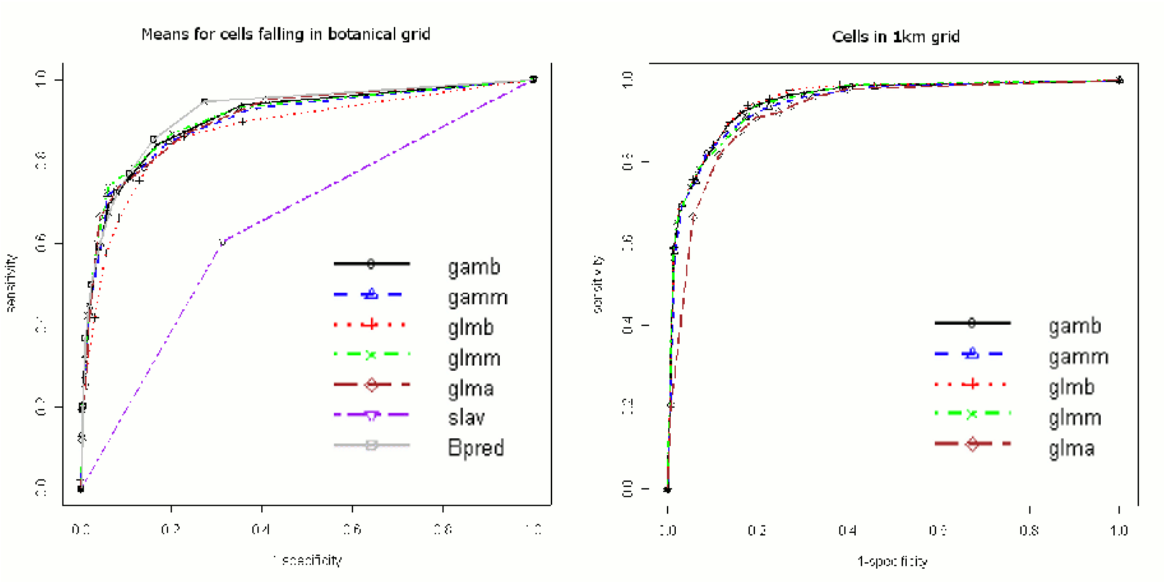


Fig.10. Comparison of models using ROC curves. The imaginary diagonal line [0,0] [1,1] means the worst ability of the model to explain observed pattern of boxelder occurrences (e.g. random pattern); the closer goes the curve to the borders of graph the better predictions model gives. For models descriptions see Tab.6.

vertically and the reversed scale of the specificity (the probability of false predicted presences) on the horizontal axis.

All models built in this study were similar in curve shape - in addition ROC statistics differ only partly (Tab.6). The Bgam model shows the best ROC statistic, which is significantly better than in the case of gam_m and glm_b. In the scale of the 1km grid only the glm_a model exhibited significant difference in prediction success. Understandably, all models were significantly better in prediction than the independent dataset derived from Slavík's map. The "True/False" plot (Fig.11), where the number of true predicted versus the number of false predicted values for 0,5 cut-off (the predicted value > 0,5 is predicted presence and < 0,5 is predicted absence) is shown, can provide more precise imagination of the difference. The

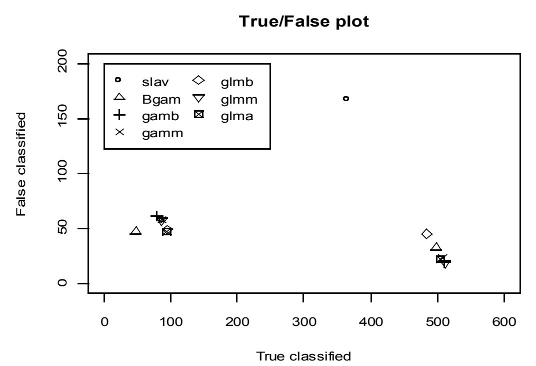


Fig.11. Number of true classified absences/ presences versus mis-classified ones is plotted. Note the single circle near legend - there is three times more false-classified presences in map from Slavik.

realized prediction is shown in the maps, where Bpred model (Fig.12) and gam_b model (Fig.13) are shown. The gam_b model was chosen for the demonstration although the glm_b model exhibited a better fit to real data, because in areas outside the boxelder‘s range the prediction from the glm_b model did not seem as precise as from gam_b. Predictive maps for all fitted models can be found at the end of appendix.

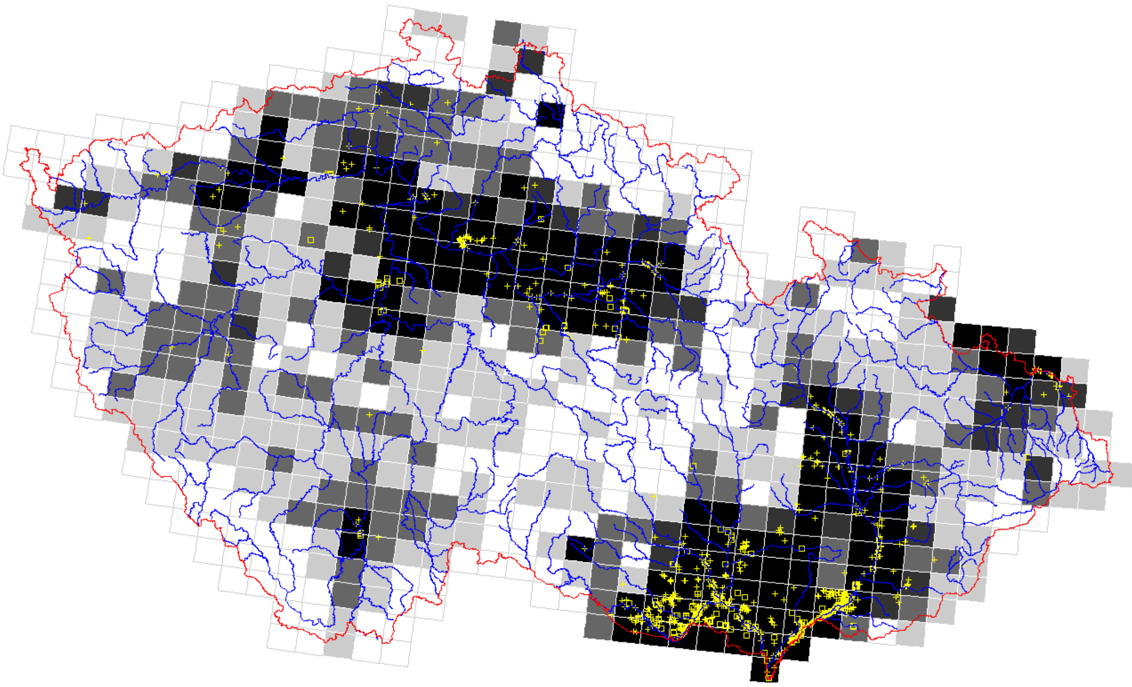


Fig.12. Prediction in the standardized mapping grid scale. The white squares means zero probability of Acer negundo wilding and the black ones the probability greater than 0,7. Whole scale is 0-0,1-0,3-0,5-0,7-1.

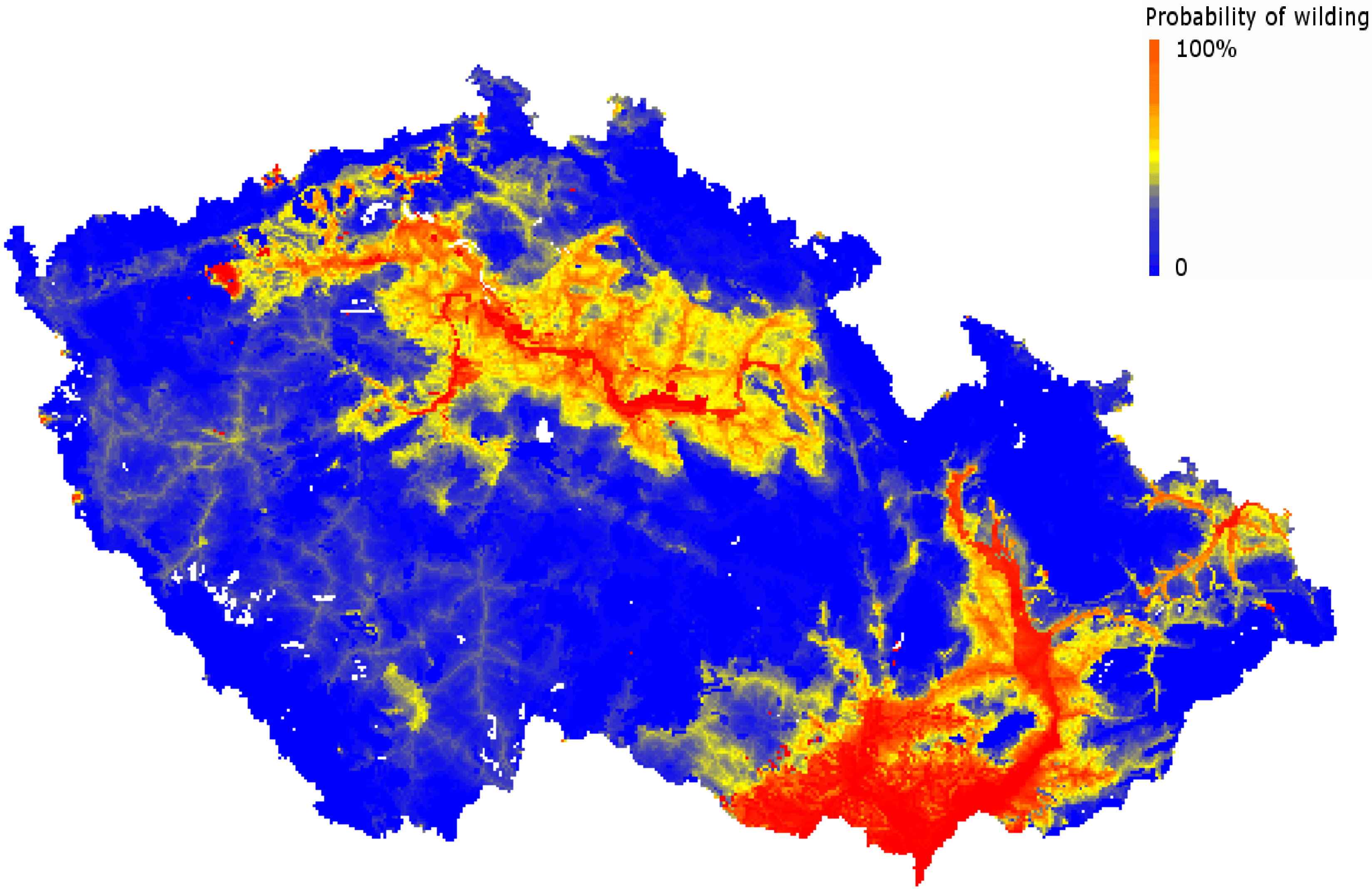


Fig.13. Prediction in a 1 km grid using the gam_b model. The blue areas are sites with zero probability of *Acer negundo* wilding, the red are sites with very high probability. The white areas are NA cells originated from levels of mpv (units of potential vegetation) which were not included in the dataset for model fitting.

4.5 Applied prediction

Predictive mapping can be realized for more, different purposes including e.g. the management planning in Protected Landscape Areas (PLA). The mean probability of *Acer negundo* wilding was calculated for each PLA in the Czech Republic as an example of the use for the methods discussed above. The probability layer for calculation was modeled by the gam_b model, but all the types of models used in this study would exhibit the same results.

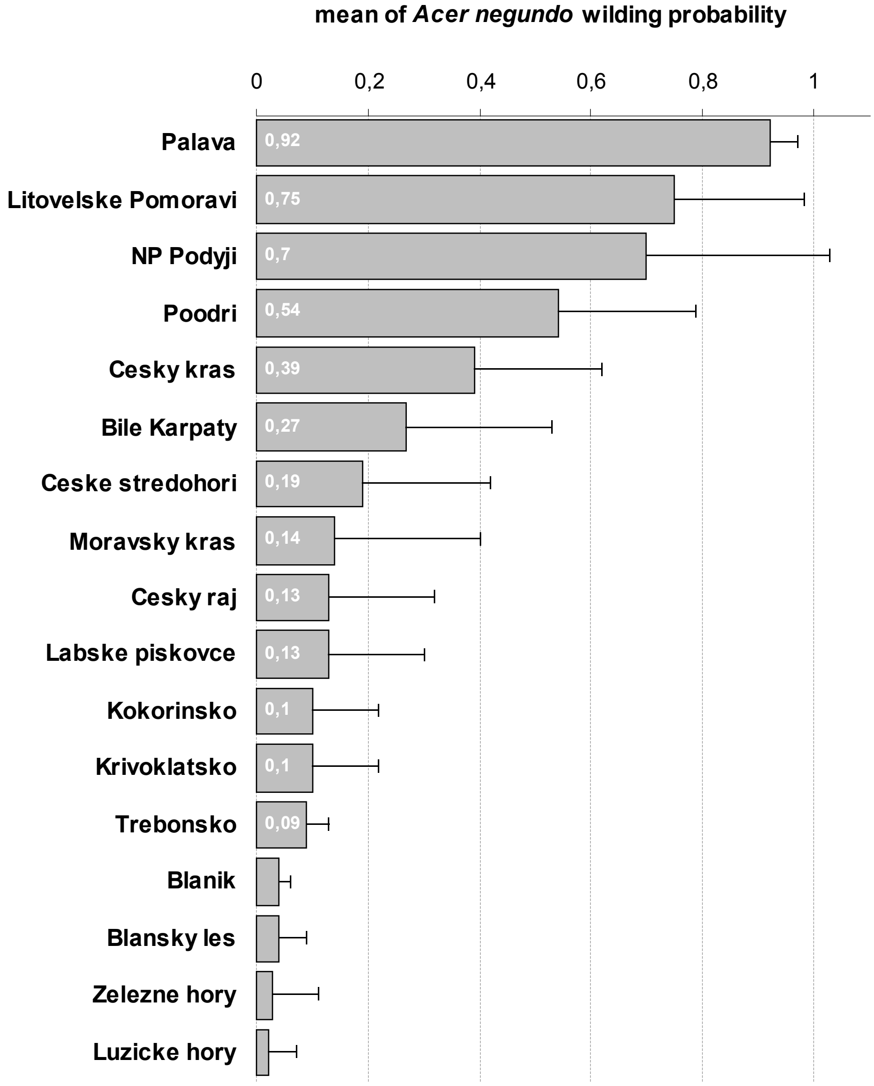


Fig.14. Probability of Acer negundo wilding in Landscape Protected areas in the Czech republic. Only averages above 0,01 are shown.

5. Discussion

5.1 Habitats

The acquired results in habitat preference of wilding *Acer negundo* agree with a few available information sources. Slavík (1972) reports naturalization in alluvial forests and riparian scrubs in the neighborhood of great rivers. The same author notices in *Flora of the Czech Republic* (vol. 5, 1997) wilding of boxelder at railway stations, abandoned courts and gardens. Such examples were documented by phytocenological releves and on this basis *Acer negundo* can be counted among weed trees. If its strong regeneration ability is considered (Maeglin et Ohmann 1973), boxelder could be posited as very problematic and reduction-resistant weed.

In the releve collection there is evidence for wilding *Acer negundo* in ruderal vegetation. This happens often in the presence of *Calamagrostis epigejos*, which is commonly regarded as a succession slowing element (Prach et al. 2001). *Acer negundo* seems to deal with such impact of competitive grasses which is possibly explained by specific bitter substances involved in its vegetative organs. These substances may be effective anti-herbivorous protection. Several *Acer negundo* mono-dominant stands were observed (and noticed in releves) in abandoned meadows and on land waiting for new urban development, often containing very few understorey species. Even though this is common in even-aged young stands, the litter of boxelder should also be investigated for any potential allelopathic substances (reported in Kolesnichenko et Spakhov 1969). The question is, in which direction the succession of such places will continue.

The new boxelder wilding habitat in the Czech Republic is reported from recent national vegetation mapping - oak-hornbeam woodlands (un. *Carpinion*). At least one occurrence of spontaneous spreading of boxelder was found in each of four phytogeographically differentiated associations from the *Carpinion* union. The horn-beam woodlands invaded by boxelder have been reported in Poland (Tokarska-Guzik 2003). The increasing number of standardized mapping grid squares with *Acer negundo* occurrence and newly invaded habitats suggest that the process of invasion is still continuing. This consequently leads to following conclusion: The predicted probabilities are in consistence with recently known boxelder wilding sites and so the sites with high probability of the boxelder wilding can be regarded as predicted correctly. Contrariwise, sites with low predicted probability are not so clearly estimated and the probability can increase, if only the early or middle phase of invasion is continuing.

5.2 Predictors

The distribution of values shows clear trends in all investigated variables. Clear positive correlation can be noted in cases of mean annual temperature and slightly in southwestness. Contrary negative correlation is evident in elevation, slope, mean annual sum of precipitation, distance to river and general in distance to man-related landscape features (in descending order: drw, dtw, dro, dci, drn). One can argue that the effect of the presence-sampling strategy can decrease averages of dro and drn and also of dci and dtw but on the contrary the most out-of-the-way areas are the most valuable for nature conservation and so the best proved by the national mapping.

While elevation has shown the best predictive power, the other predictors were found redundant for model with only one exception: potential vegetation units. This is natural because all features investigated for distance to *Acer negundo* wilding sites are landscape units situated in low altitude. Moreover, the netting effect in smaller scale, common to all these landscape features, decreases their effects in country level modeling. Such predictors are much better when in greater scale are used (Urban 2003). The surprising thing was that distance to a river, although it is clear relevant predictor (see natural habitat for *Acer negundo* in section 1.4), was cut out from all built models. This can be explained through the use of potential vegetation units, of which the alluvial forests are more acurate predictors. While the ***drv*** increases linearly with distance to a river, the vegetations types reflect alluvium borders, which is naturally much better for modeling the probability of any alluvial plant occurrence. The second reason could be the scale of the 1 km grid, while more than half occurrences were found closer than 1km. Here can also be noted that the distance to a river can be used, if spreading rather than distribution is modeled.

The next important problem in deriving “distance from“ variables is fault in data from marginal border-close areas. Features such as railways and roads can be closer to sites behind the border, but in our data layers this information is not included.

The role of climatic variables is also complicated, because of altitude-dependent interpolation used for their calculation from individual climatic stations. For example, the heat-islands around cities are ignored in such layers, even though they are often stressed in plant invasion literature (Gilbert 1989). More precise temperature layer could be obtained from remote sensing, but such data were not available. Lehmann et al. (2002) show that the seasonal characteristics and hydrological related variables can contribute to more precise prediction due to the closer relationships to plant physiology.

Because of correlation among variables it is not legitimate to interpret results of the t-test as an evidence of causality, but they are sufficient for imagining the relationship between variable values and occurrence of wilding *Acer negundo*.

5.3 Model types

All the models exhibit similar accuracy of prediction when the same dataset was used for model fitting and model evaluation. But if predicted surfaces are compared larger differences are evident. The GLM based models overestimate the probability of *Acer negundo* wilding for some rare types of geology and therefore islands of very high probability occur in the middle of large zero areas (see glm_b model map in appendix). Other differences originate in other predictor selection. Before realizing a predictive map using this study, a calibration of the selected model should be performed. Because it is a time-consuming process it was not included here, but together with appropriately adjusted selection of pseudoabsences such practice can improve the predictive surface layout (Guisan and Zimmerman 2000).

GRASP is one of predictive systems based on logistic regression and thus needs information about absence sites for successful prediction. Sampling pseudo-absences can bring uncertainty into estimated values. Another method can be used for the same goal without absence data: ENFA (Hirzel et al. 2002) based on habitat suitability modeling. However Zaniewski et al. (2002) show GAM model with pseudo-absences which gives closer result to GAM with true absences than the ENFA approach.

5.4 Prediction

The resulting prediction maps shows the climatic equivalent of recently boxelder wilding sites rather than all potential sites of boxelder occurrence. As mentioned above the oak-hornbeam woodlands are newly invaded by *Acer negundo* and this can continue into a more wide-spread invasion. The large ecological valence of this tree enables such a process namely if this species was observed in 2300 m altitude (Maeglin et Ohmann 1973) in natural areal (of course in closed riparian bush wood).

The white cells in predictive maps are caused by missing levels for mpv factor in the dataset used for model fitting. E.g. category 49 - complex of submontane *Pinus rotundata* and *P. sylvestris* mires is small-area unit from South Bohemia and was not sampled in random pseudoabsence sampling. To avoid this mistake the stratified random sampling should be used for pseudo-absence data sampling (Urban 2003).

6. Conclusions

6.1 Habitats

Before known habitats of *Acer negundo* wilding were confirmed and the new ones were observed. This can indicate the invasion process is not yet finished in the Czech Republic.

6.2 Predictors

Elevation can explain most of the boxelder wilding site distribution. It is not a primary predictor, but it is highly correlated to the direct predictor - mean annual temperature. In our case can be achieved much more accurate results, because of the higher accuracy of our elevation data, than our temperature data.

Other used predictors can contribute to precise prediction, but also can bring more bias. The selection of predictors group is target-area specific and such should be done with respect to the use of results. Which of variables are used alternate with different types of model technique (GAM versus GLM in this study).

6.3 Model types

GLM and GAM were used in this study and GAM seems be more useful for predicting outside the range of presences dataset. When compared goodness of fit among these techniques with the same dataset which was used for model fitting no significant difference was observed.

6.4 Scales

Also different scale was used for prediction (with different approach to variable derivation),but conversion of values predicted in 1km grid scale to 10km grid scale did not brought any better results. Conclusion can be that the prediction should be undertaken in each resolution in which predicted data will be used.

7. References

Beck J.R., Schultz E.K., 1986. The use of Relative Operating Characteristic (ROC) Curves in Test Performance Evaluation, Arch Pathol Lab Med, 110: 13-20

Brink, V.C., 1954. Survival of plants under flood in the lower Fraser River Valley, British Columbia, Ecology, 35: 94-95

Brown, J.H., Mehlman,D.H. & Stevens, G.C., 1995. Spatial variation in abundance, Ecological Modeling, 76: 2028-2043

Busby,J.R., 1986. A biogeographical analysis of Nothofagus cunninghamii (Hook.) Oerst. in southeastern Australia, Australian Journal of Ecology, 11: 1-7

Busby,J.R., 1991. BIOCLIM - A Bioclimatic Analysis and Prediction System. In: Margules, C.R.& M.P. Austin (eds.), Nature Conservation: Cost Effective Biological Surveys and Data Analysis, CSIRO, Canberra

Collingham, Y.C., Hill, M.O. & Huntley, B., 1996. The migration of sessile organisms: a simulation model with measurable parameters, Journal of Vegetation Science, 7: 831-846

Cronk, Q.C.B., Fuller, J.L., 1995. . In: , Plant Invaders, Chapman & Hall, UK

Dettki, H., Loefstrand, R. & Edenius, L., 2003. Modeling habitat suitability for moose in coastal northern Sweden: empirical vs. process-oriented approaches, AMBIO, 32 (8): 549-556

di Castri, F, 1990. On invading species and invaded ecosystems: the interplay of historical chance and biological necessity. In: di Castri, F., Hansen, A., J., Debussche, M. Biological Invasions in Europe and the Mediterranean Basin, Biological Invasions in Europe and the Mediterranean Basin, Kluwer Academic Publischers, The Netherlands

Dollar, K.E., Pallardy, S.G., Garrett, H.G., 1992. Composition and environment of floodplain forest of northern Missouri, Cam. J. For. Res, 22: 1343-1350

Eilth, J. and Burgman, M., 2001. Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. In: Scott J.M., Heglund P.J., Morrison M., Raphael M., Haufler J., Wall B., Samson F., Predicting Plant, Animal and Fungi Occurrences: Issues of Scale and Accuracy, sland Press, Clovelo (CA)

Franklin, J., 1995. Predictive vegetation mapping: geographical modeling of biospatial patterns in relation to environmental gradients, Prog. Phys. Geogr, 19: 474-499

Gilbert, O., 1989. The ecology of urban habitats., Chapman and Hall, London:

Guisan, A., Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology, Ecological Modeling, 135: 147-186

Guth, J., 2002. Metodiky mapování biotopů soustavy NATURA 2000 a SMARAGD (metodiky podrobného a kontextového mapování) [Methodology for NATURA 2000 a SMARAGD habitats mapping], AOPK ČR, Praha: 37 pp.

Hadincová, V., Dobrý, J., Hanzalyová, D., Härtel, H., Herban, T., Krahulec, F., Kyncl, J., Moravcová, L., Šmilauer, P., Šmilauerová, M., 1997. nvazní druh Pinus strobus v Labských pískovcích, Zprávy Čes. Bot. Společ. Praha, 32, Mater. 14: 63-79

Hanley J.A., McNeil B.J., 1983. A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases., Radiology, 148: 839-43

Hastie, T., Tibschirani, R., 1986. Generalized additive models, Statistical Science, 3 (1): 297-318

Hastings, A, 1996. Model of spatial spread: a synthesis, Biological Conservation, 78: 143-148

Hill, M.O., Bunge, M.O., et Shaw, M.W., 1975. Indicator species analysis, a divisive polythetic method of classification, and its application to a survey of native pinewoods in Scotland, Journal of ecology, 63: 597-613

Hirzel A.H., Hausser J., Chessel D. & Perrin N, 2002. Ecological-niche factor analysis: How to compute habitat- suitability maps without absence data?, Ecology, 83: 2027-2036

Hirzel, A., Hausser, J., Perrin, N., 2001. Biomapper 2.0. Laboratory for Conservation Biology, Institute of Ecology, University of Lausanne: URL: http://www.unil.ch/biomapper

Hutchinson, MF., 1989. A new method for gridding elevation and stream line data with automatic removal of pits, Journal of Hydrology, 106: 21-23

Chytrý, M. , Rafajová, M., 2003. Czech National Phytosociological Database: basic statistic of available vegetation-plot data., Preslia, Praha, 75: 1-15

Chytrý, M., Kučera, T. and Kočí, M., 2001. Katalog biotopů České republiky, Agentura ochrany přírody a krajiny ČR, Praha: 304 p.

Jones, P.G., Gladkov, A., 1999. FloraMap Version 1. A computer tool for predicting the distribution of plants and other organisms in the wild, CD-ROM and Manual, Centro Internacional de Agricultura Tropical, Cali Colombia: 99 p.

Kolesnichenko, M.V. and Spakhov, Y.M., 1969. , Fiziol Rast, 16: 1074

Komissarov, A.D., 1964. Biologicheskie osnovy razmnozheniya rastenii cherenkami [Biological basis for the propagation of woody plants by cutting], Izdaťelstvo Lesnaya promyshlennost, Moskua, [Transl. Israel Prog. Sci. Transl. 1968. Jerusalem.]: 250 p.

Kowarik, I., 1995. Time lag in biological invasions. In: Pyšek, P., Prach, K., Wade, M., Plant invasions - General aspects and special problems, SPB Academic Publising, Amterdam, The Netherlands

Kubát, K., Hrouda, L., Chrtek, J. jun., Kaplan, Z., Kirschner, J., 2002. Klíč ke květeně České republiky, Academia, Praha: 927 p.

Lehmann, A., Overton, J.McC., Leathwick, J.R., . GRASP: Generalized Regresion Analysis and Spetial Prediction, Ecological Modeling, 157: 189-207

Lehmann, A., Overton, J.McC., Leathwick, J.R., . GRASP: Generalized Regresion Analysis and Spetial Prediction, Ecological Modeling, 157: 189-207

Lek S., Delacoste, M., Baran, P., Dimopoulos I., Lauga J. & Aulanier S., 1996. Application of neural networks to modeling non-linear relationships in ecology, Ecological Modeling, 90: 39-52

Macdonald, I.A.W., 1990. Strategies for limiting the invasion of protected areas by introduced organisms, Monogr. Syst. Bot. Missouri Bot. Gard, 32: 1-11

Maeglin, R.R., Ohmann, L.F., 1973. Boxelder (Acer negundo): A review and commentary, Bulletin of the Torrey botanical club, vol. 100, n. 6: 357-363

Mastrorillo, S., Lek, S., Dauba, F. & Belaud, A., 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river, Freshwater

Biology, 38: 237-246
Neuhäuselová, Z. [ed.], 1998. Map of potential natureal vegetation of the Czech republic, Academia, Praha: 341 p.
Noble, I.R., 1989. Attributes of invaders and the invading process: terrestrial and vascular plants, , : 301-313
Patthey, P., 2003. Habitat and corridor selection of an expanding red deer (Cervus elaphus) population, PhD. Thesis, Institut of Ecology: Univesity of Lausanne
Peterson AT and Cohoon, KP, 1999. Sensitivity of distributional prediction algorithms to geographic data completeness, Ecological Modeling, 117: 159-164
Peterson, A.T., 2001. Predicting species geographic distributions based on ecological niche modeling, The Condor, 103: 599-605
Prach K., Pyšek P., Bastl, M., 2001. Spontaneous vegetation succession in human-disturbed habitats: A pattern across seres, Applied Vegetation Science, 4: 83-88
Pyšek P., Sádlo J. & Mandák B., 2002. Catalogue of alien plants of the Czech Republic, Preslia, Praha, 74: 97–186
Pyšek, P., 2001. Past and future of predictions in plant invasions: a field test by time, Diversity and Distributions, 7: 145–151
Pyšek, P., Sádlo, J. & Mandák, B., 2003. Alien flora of the Czech republic, its composition, structure and history. In: Child, L., Brock, J.H., Brundu, G., Prach, K., Pyšek, P., Wade, P.M. & Williamson, M., Plant invasions: Ecological threads and mangement solutions, Buchuys Publishers, Leiden
Richardson, D.M., Macdonald, I.A.W., Fosyth, G.G., 1989. Reductions in plant in plant species richness under stands of alien trees and shrubs in the Fynbos biome, South African Forestry Journal, 149: 1-8
Slavík, B., 1972. Naturalizace javoru jasanolistého (Acer negundo L.) v Československu, Acta musei silesiae, series dendrologia,2: 154-162
Slavík, B., 1997. Fytogeografické syntézy ČR 3., [Phytogeographycal synthesis of Czeck Republic, volume 3], Praha: 160 p.
Slavík, B., 1997. Květena České republiky. 5. [Flora of the Czech Republic], Academia, Praha: 568
Stockwell, D.R.B., Peters, D., 1999. The GARP modeling system: problems and solutions to automated spatial prediction, International Journal of Geographic Information Science, 13: 143-158
ter Braak, C. J. F. and Šmilauer, P, 2002. CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (version 4.5)., Microcomputer Power, Ithaca NY, USA: 500 pp.
ter Braak, C. J. F., 1986. Canonical correspondence analysis: a new eigenvector method for multivariate direct gradient analysis., Ecology, 67: 1167-1179
Tokarska-Guzik, B., 2003. The expansion of some alien plant species (neophytes) in Poland. In: Child, L., Brock, J.H., Brundu, G., Prach, K., Pyšek, P., Wade, P.M. & Williamson, M., Plant invasions: Ecological threads and mangement solutions, ,
Urban, D.L., 2003. Lecture notes for Spatial Analysis in Ecology course., Landscape Ecology Laboratory, Duke University, Durham, USA: http://www.env.duke.edu/landscape/classes/env352/sa_syl.html

Vitousek, P.M., D'Antonio, C.M., Loope, L.L., Rejmanek, M. & Westbrooks, R., 1997. Introduced species: a significant component of human caused global change, New Zealand Journal of Ecology, 21: 1-16

Wadsworth, R. A., Y. C. Collingham, S. G. Willis, B. Huntley and P.E. Hulme, 2000. Simulating the spread and management of alien riparian weeds: are they out of control?, Journal of Applied Ecology, 37: 28-38

Williamson, M.H., Fitter, A., 1996. The characters of succesful invaders, Biological Conservation, 78: 163-170

Zalba, M.S., Sonaglioni, M.I., Compagnoni A.C., Belenguer, C.J., 2000. Using a habitat model to assess the risk of invasion by an exotic plant, Biological Conservation, 93: 203-208

Zaniewski, A.E., Lehmann, A., Overton, J.McC., 2002. Predicting species spatial distribution using presence-only data: a case study of native New Zealand ferns, Ecological Modeling, 157: 261-280

8. Related www sites

8.1 Organisations:

Czech National Phytosociological Database,	http://www.sci.muni.cz/botany/database.htm
AOPK CR,	http://www.nature.cz/
Natura2000,	http://www.natura2000.cz/
ArcData, Praha, DEM of the Czech Republic	http://www.arcdata.cz/data/dmr

8.2 Software:

The R Project for Statistical Computing	http://www.r-project.org/
GRASS GIS	http://grass.itc.it/index.php
Analyse it (ROC analysis)	http://www.analyse-it.com/default.asp

9. Appendix

Tab.7. Describtion of the categories in potencial vegetation map (Neuhäuselová 1998)

n.	description of potential vegetation type
1	Bird cherry-ash woodland (Pruno-Fraxinetum Oberdorfer 1953), partly in complex with adler carrs (Alnion glutinosae Malcuit 1929).
2	Bird cherry-oak and -adler woodland (spol. Quercus robur-Padus avium, spol. Alnus glutinosa-Padus avium) with Carex brizoides, partly in complex with adler carrs (Carici elongatae-Alnetum Schwickerath 1933) and reed swamps and tall-sedge communities (Phragmito-Magnocaricetea)
3	Spruce-alder woodland (Piceo-Alnetum Rubner ex Oberdorfer 1957).
4	Poplar-pedunculate oak woodland (Querco-Populetum Neuhäuslová-Novotná 1965), partly in complex with elm-pedunculate oak woodland (Querco-Ulmetum Issler 1926).
5	Elm-pedunculate oak woodland (Querco-Ulmetum Issler 1926).
6	Panonian elm-ash woodland (Fraxino pannonicae-Ulmetum Soó in Aszód 1936 corr. Soó 1963) partly in complex with poplar-ash woodland (Fraxino-Populetum Jurko 1958).
7	Oak-hornbeam woodland with Melampyrum nemorosum (Melampyro nemorosi-Carpinetum Passarge 1957).
8	Lime-oak woodland with Betula pendula (Tilio-Betuletum Passarge 1957).
9	Panonian oak-hornbeam woodland with Primula veris (Primulo veris-Carpinetum Neuhäusl et Neuhäuslová ex Neuhäuslová-Novotná 1964).
10	Carpathian oak-hornbeam woodland with Carex pilosa (Carici pilosae-Carpinetum Neuhäusl et Neuhäuslová 1964).
11	Lime-rich oak-hornbeam woodland (Tilio-Carpinetum sensu Traczyk 1962).
12	Lime-pedunculate oak woodland with Stellaria holostea (Stellario-Tilietum Moravec 1964).
13	Scree and ravine woodland of colline to montane sites. (Aceri- Carpinetum Klika 1941, Lunario-Aceretum Schlüter in Grüneberg et Schlýter 1957, Mercuriali-Fraxinetum [Klika 1942] Husová 1982, Scolopendrio-Fraxinetum Schwickerath 1938).
14	Lime-beech woodland with Tilia platyphylos (Tilio platyphylli-Fagetum Klika 1939).
15	Lime-beech woodland with Tilia cordata (Tilio cordatae-Fagetum Mráz 196O em. Moravec 1977).
16	Beech woodland with Melica uniflora (Melico-Fagetum Seibert 1954).
17	Sedge-rich beech woodland with Carex pilosa (Carici pilosae-Fagetum Oberdorfer 1957).
18	Beech woodland with Dentaria enneaphyllos (Dentario enneaphylli-Fagetum Oberdorfer ex W. et A. Matuszkiewicz 196O).

n.	description of potential vegetation type
19	Carpathian beech woodland with Dentaria glandulosa (Dentario glandulosae-Fagetum Matuszkiewicz ex Guzikowa et Kornas 1969).
20	Beech woodland with Festuca altissima (Festuco altissimae-Fagetum Schlüter in Grüneberg et Schlüter 1957).
21	Beech woodland with Viola reichenbachiana (Violo reichenbachianae-Fagetum Moravec 1979).
22	Beech woodland with Cephalanthera sp. (Cephalanthero-Fagetum Oberdorfer 1957).
23	Silver fir woodland with Sanicula europea (Saniculo europaeae-Abietetum Husová [1968] nom. nov.).
24	Woodrush-beech woodland (Luzulo-Fagetum Meusel 1937).
25	Spruce-beech woodland (Calamagrostio villosae-Fagetum Mikyška 1972).
26	Waterlogged pedunculate oak-beech woodland with Carex brizoides (Carici brizoidis-Quercetum Neuhäusl in Mikyška et al. 1968)
27	Silver fir woodland with Deschampsia flexuosa (Deschampsio flexuosae-Abietetum Husová 1968).
28	Oak woodland with Lathyrus versicolor and/or Buglossoides purpurocoerulea (Lathyro versicoloris- Quercetum pubescentis Klika [1928] 1932, Torilido-Quercetum Blažková 1997).
29	Oak woodland with Prunus mahaleb and/or Cornus mas (Pruno mahaleb-Quercetum pubescentis Jakucs et Fekete 1957, Corno-Quercetum Máthé et Kovács 1962).
30	Undeterminated basiphilous thermophilous oak woodland (Brachypodio pinnati-Quercetum Klika 1953 nom. inv. aj.).
31	Oak woodland on loess with Quercus petraea, Q. pubescens, Q. robur (Quercetum pubescenti-roboris [Zólyomi 1957] Michalko et Džatko 1965).
32	Subcontinental pedunculate oak woodland with Carex fritschii (Carici fritschii-Quercetum roboris Chytrý et Horák 1997).
33	Oak woodland with Potentilla alba (Potentillo albae-Quercetum Libbert 1933).
34	Oak woodland with Sorbus torminalis and Vincetoxicum hirundinaria (Sorbo torminalis-Quercetum Svoboda ex Blažková 1962).
35	Oak woodland with Asplenium cuneifolium on serpentine substrate (Asplenio cuneifolii-Quercetum petraeae).
36	Woodrush oak and/or silver fir-oak woodland (Luzulo albidae-Quercetum petraeae Hilitzer 1932, Abieti-Quercetum Mráz 1959).
37	Oak woodlalnd with Molinia coerulea (Molinio arundinaceae-Quercetum Neuhäusl et Neuhäuslová-Novotná 1967).
38	Pine-oak woodland with Vaccinio vitis-idaea (Vaccinio vitis-idaeae-Quercetum Oberdorfer 1957).

n.	description of potential vegetation type
39	Pine-oak woodland with Festuca ovina (Festuco ovinae-Quercetum roboris sensu F. Őmarda 1961).
40	Pine woodland with Thlaspy montanum on serpentine substrate (Thlaspio montani-Pinetum sylvestris Chytrý in Chytrý et Vicherek 1996).
41	(Sub)montane spruce-pine and spruse woodland on stony substrate (Betulo carpaticae-Pinetum Mikyška 1970, Anastrepto-Piceetum Stöcker 1967).
42	Other acilophilous pine woodland (Dicrano-Pinion [Libbert 1933] Matuszkiewicz 1962 excl. Betulo carpaticae-Pinetum Mikyška 197O, Vaccinio uliginosi-Pinetum sylvestris Kleist 1929).
43	Spruce woodland with Calamagrostis villosa (Calamagrostio villosae-Piceetum Hartmann in Hartmann et Jahn 1967).
44	Waterlogged spruce woodland with Bazzania trilobata (Mastigobryo-Piceetum [Schmid et Gaisberg 1936] Braun-Blanquet, Sissingh et Vlieger 1939), partly in complex with Sphagnum-rich spruce woodland (Sphagno-Piceetum sensu Sofron 1981).
45	Spruce woodland with Athyrium distentifolium (Athyrio alpestris-Piceetum [Hartmann 1959] Hartmann et Jahn 1967).
46	Complex of Pinus mughi communities and alpine vegetation(Pinion mughi Pawlowski in Pawlowski, Sokolowski et Wallisch 1928)Pinus sylvestris-mires (Pino rotundatae-Sphagnetum Kästner et Fl÷ssner 1933 corr. Neuhäusl 1969, Eriophoro vaginati-Pinetum sylvestris Hueck 1931 em. Neuhäusl 1984), Vaccinio uliginosi-Pinetum sylvestris Kleist 1929).47 Complex of sedge and sedge-moss communities on minerotrophic mires (Caricetalia fuscae Koch 1926).
48	Complex of sedge-Sphagnum communities on minerotrophic mires (Scheuchzerietalia palustris Nordhagen 1936 excl. Leuko-Scheuchzerion palustris Nordhagen 1943).
49	Complex of submontane Pinus rotundata- and
50	Complex of montane raised bogs (Sphagnetalia medii Kästner et Flössner 1933 excl. sub 49), partly with Pinus mugo agg. and/or Sphagnum-rich spruce woodland (Sphagno-Piceetum sensu Sofron 1981).
51	Comlex of succesional stages on anthropogenic sites (open-cast coal mines etc.)
52	Water (damms, flooded areas)

Tab.8. Describtion of the categories in geological theme

n.	describtion
10	monotonous serie of Moldanubicum (mica shist-gneiss, paragneiss, migmatite)
11	variegated serie of Moldanubicum (mica shist-gneiss, paragneiss-migmatite with intercalated beds of limestone, erlan, quartzite, graphite and amphibolite)
12	Proterozoic Assynt folded rocks, with a differently strong Variscan reprocess (shale, phyllite, mica shist-paragneiss)
13	orhogneiss, granulite and diatexite in Moldanubicum and Proterozoic
14	ultrabasic rock in Moldanubicum and Proterozoic
15	Assynt granitoit (granite, granodiorite)
16	Palaeozoic folded and metamorphosed rock (phyllite, mica shist)
17	Palaeozoic folded and unmetamorphosed rock (shale, wacke, quartzite, limestone)
18	Proterozoic-Palaeozoic volcanic rock partly metamorphosed(amphibolite, diabase, melaphyre, porphyry)
19	granite (granite series)
20	granodiorite-diorite (tonalite series)
21	dark granodiorite, syenite (durbachite series)
22	diorite and gabbro, Assynt and Variscan
23	Mesozoic rock Alpine folded (sandstone, shale)
24	Tertiary rock Alpine folded (sandstone, shale)
25	Anthracolithic rock (sandstone, conglomerate, claystone)
26	Mesozoic rock (sandstone, claystone)
27	Tertiary rock (sand, clay)
28	Teriary volcanic rock (basalt, phonolit, tuff)
29	Quaternary (loam, loess, sands, gravel/broken stone)

Tab.9. Names of species used in ordination diagram.

abbr.	**Latin name**
AcerNegu	*Acer negundo*
AegoPoda	*Aegopodium podagraria*
AchiMill	*Achillea millefolium* agg.
AlnuGlut	*Alnus glutinosa*
AlopAequ	*Alopecurus aequalis*
AlopPrat	*Alopecurus pratensis*
ArrhElat	*Arrhenatherum elatius*
ArteVulg	*Artemisia vulgaris*
BideTrip	*Bidens tripartitus*
BracSylv	*Brachypodium sylvaticum*
CircLute	*Circaea lutetiana*
CirsArve	*Cirsium arvense*
CypeFusc	*Cyperus fuscus*
ElymRepe	*Elymus repens*
FestGiga	*Festuca gigantea*
FraxExce	*Fraxinus excelsior*
GeumUrba	*Geum urbanum*
GlecHede	*Glechoma hederacea*
LamiMacu	*Lamium maculatum*
LysiNUmm	*Lysimachia nummularia*
LythSali	*Lythrum salicaria*
OenaAqua	*Oenanthe aquatica*
QuerPetr	*Quercus petraea* agg.
QuerRobu	*Quercus robur*
SaliAlba	*Salix alba*
TaraSect	*Taraxacum sect. Ruderalia*
TiliCord	*Tilia cordata*

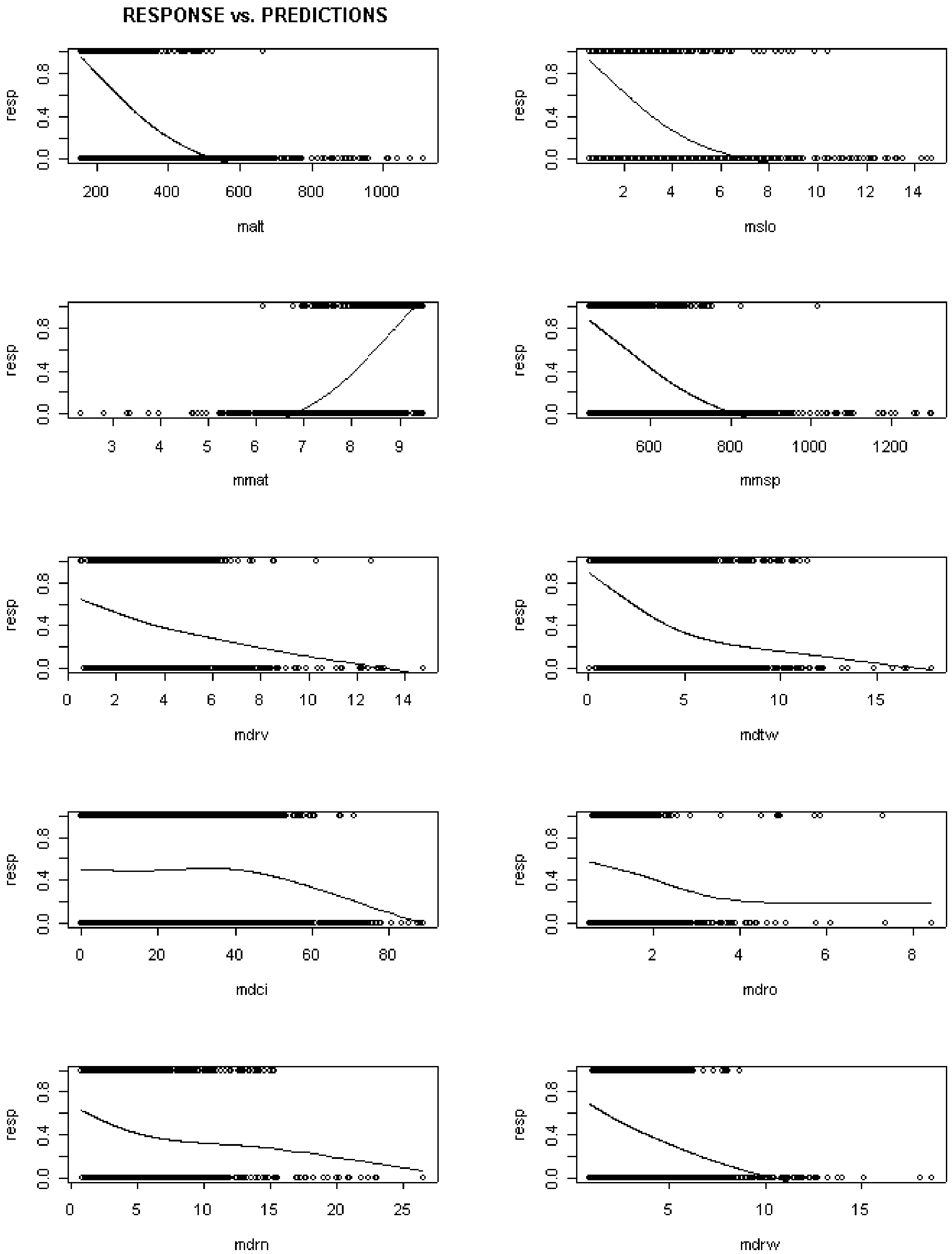


Fig.15. **Response versus predictors** are plotted and the smoothing function 'spline' is applied (4 degree of freedom) to the relationships between the responses and predictors. The names of variables are listed in Tab.2.

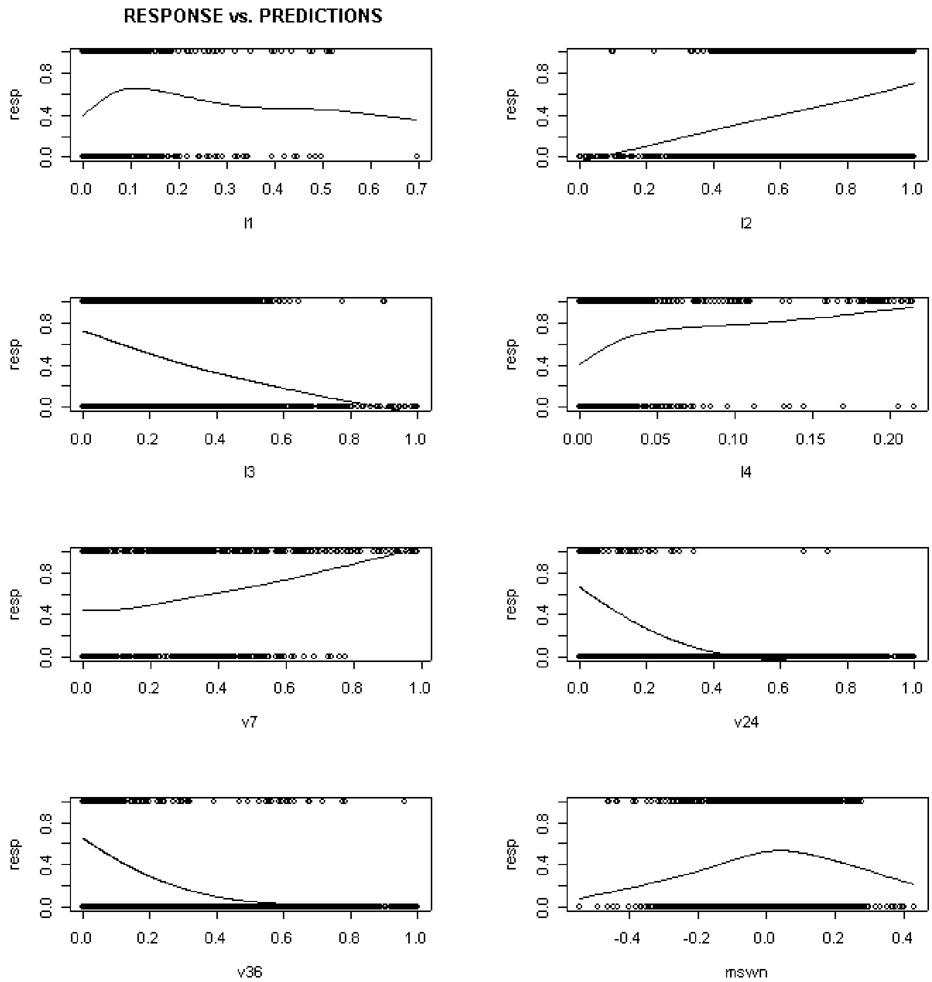
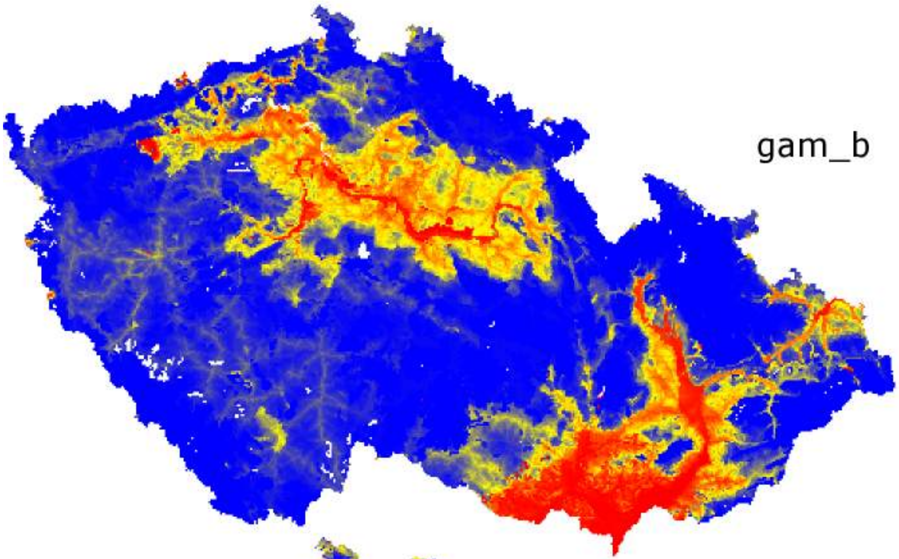
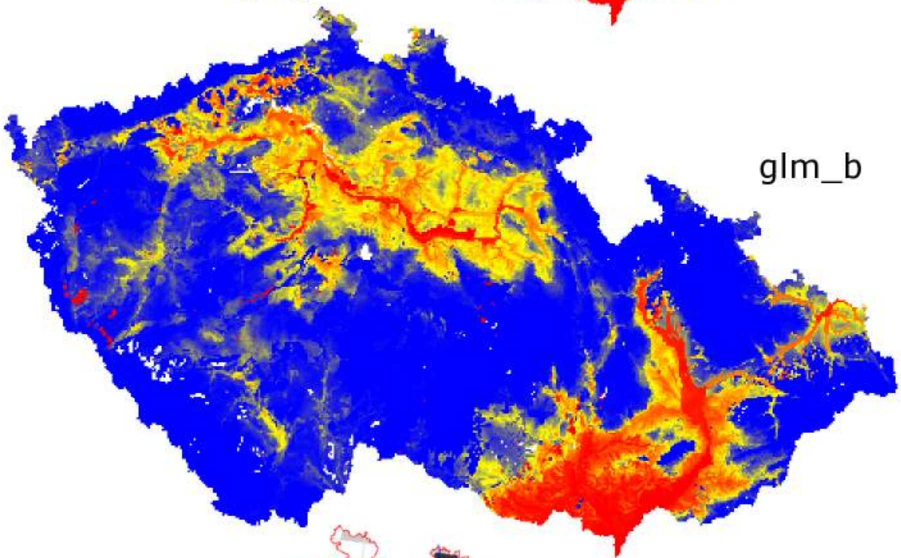
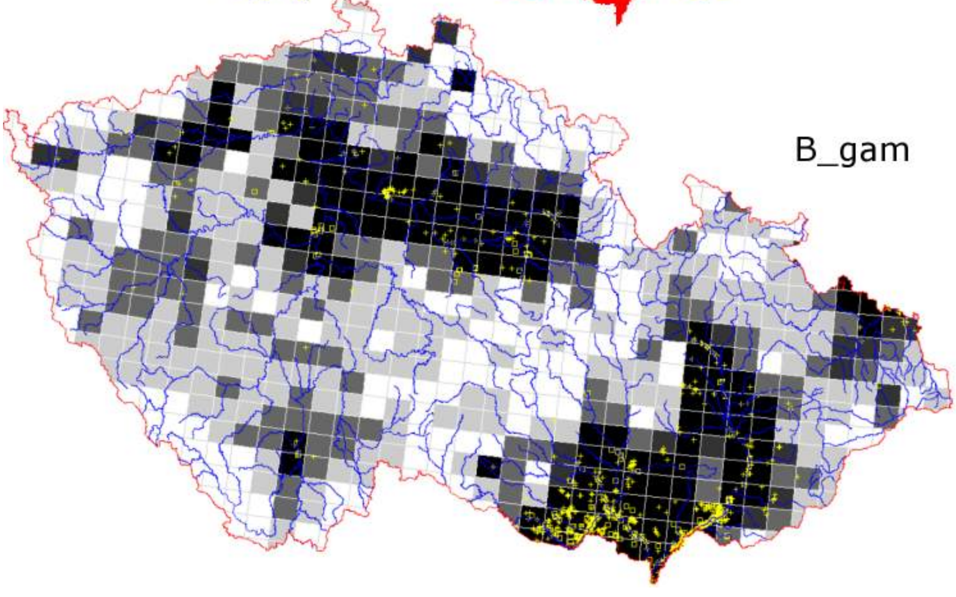


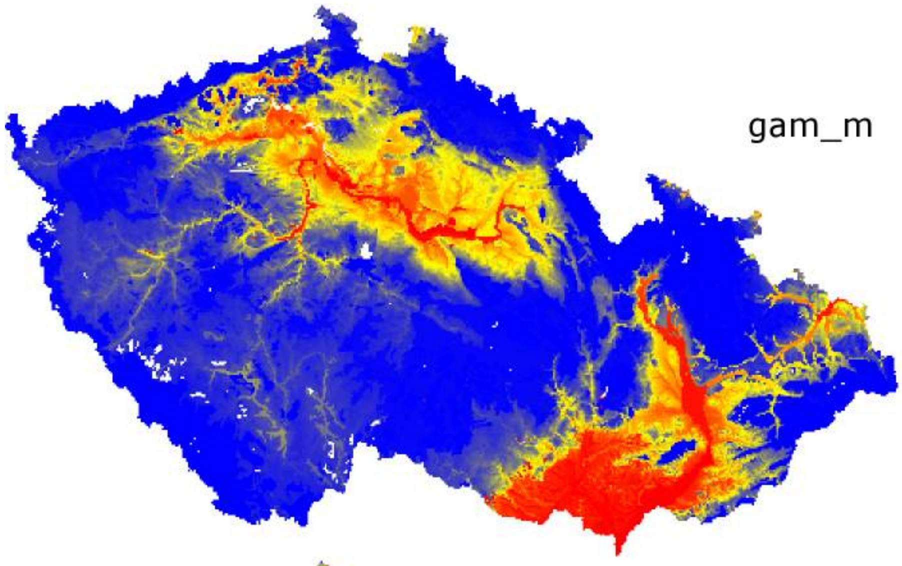
Fig.16. Response versus predictors are plotted and the smoothing function 'spline' is applied (4 degree of freedom) to the relationships between the responses and predictors. l1-l4 represent lcv factor levels, v7, v24, v36 are selected levels from map of potential elevation. The names of variables are listed in Tab.2.

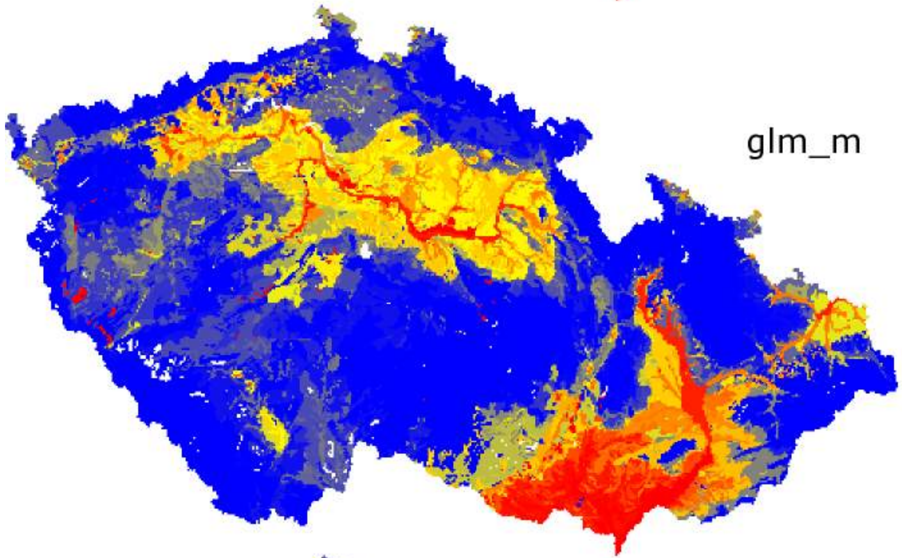
Fig.17. Correlation matrix of predictors in the 1km grid scale. Coefficient of correlation is shown in left upper corner of each graph. The names of variables are listed in Tab.2.

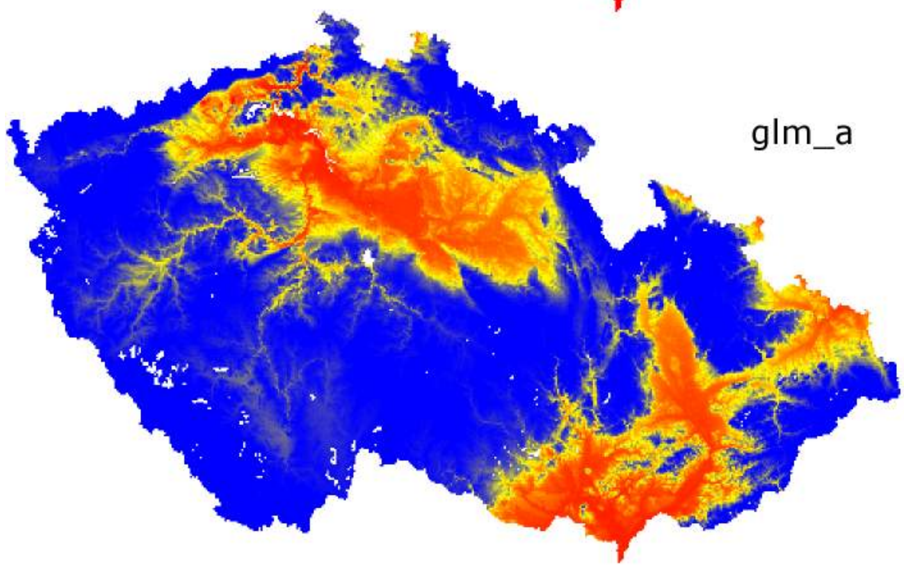

gam_b


glm_b


B_gam


gam_m


glm_m


glm_a