UNIVERSITY OF NOVA GORICA

GRADUATE SCHOOL

# DISTRIBUTION PATTERNS OF THE INVASIVE SPECIES *ROBINIA PSEUDACACIA* IN NORTHEAST SLOVENIA

MASTER'S THESIS

**Daniela Alexandra Teixeira da Costa Ribeiro**

Mentor: doc. dr. Andraž Čarni

Nova Gorica, 2011

UNIVERZA V NOVI GORICI
FAKULTETA ZA PODIPLOMSKI ŠTUDIJ

# VZORCI RAZŠIRJENOSTI INVAZIVNE VRSTE *ROBINIA PSEUDACACIA* V SEVEROVZHODNI SLOVENIJI

MAGISTRSKO DELO

**Daniela Alexandra Teixeira da Costa Ribeiro**

Mentor: doc. dr. Andraž Čarni

Nova Gorica, 2011

This master thesis was prepared at Graduate School of the Faculty of Environmental Sciences, University of Nova Gorica.

Senate from Faculty of Environmental Sciences on 19 of January 2011 named doc. dr. Andraž Čarni as mentor of this master thesis and the three following members for the Evaluation Committee:

- prof. dr. Marko Debeljak, Institute Jožef Stefan, Ljubljana
- doc. dr. Damijan Denac, National Institute of Biology, Ljubljana
- prof. dr. Eduard Hochbichler, Institute of Silviculture, University of Natural Resources and Life Sciences, Vienna.

## Abstract

*Robinia pseudacacia* L. was introduced into Europe at the beginning of the 17th century and is now considered to be an invasive species in many European countries also in Slovenia. Our study area was located in North-eastern Slovenia, within the Prekmurje region. The goal of our study was to find explanations for the current occurrence pattern of the species in that location. For this we aimed to find out which are the variables that influence the current occurrence of the species and to predict its spatial distribution for three testing sites, situated outside the training site. Areas dominated by *R. pseudacacia* have been mapped in the field in the lowland area of Prekmuje region, across a sample plot of 12km$^2$ (training site) in 2009. We analyzed potential explanatory variables that can influence the distribution of the species within the region: distance to the nearest road, distance to the nearest water body, elevation, land use, soil type and soil quality. We performed a spatial randomized sampling technique stratified for prevalence on the resulting maps in order to collect observations on the relationship between *R. pseudacacia* presence and the potential influencing factors. The statistical relationships were then established by a generalized linear model (GLM). The final model was used to predict the spatial distribution of the species for the three testing sites.

*R. pseudacacia* was found to occur mostly in parcels designated as meadows and pastures. Distance from the road network seems to facilitate the occurrence of the species to a certain degree. Distance from water bodies seems to decrease *R. pseudacacia* presence. We did not find a significant relationship between elevation and species presence, this factor apparently does not influence the distribution of the species in this region.

*R. pseudacacia* expands naturally but it is also being planted by farmers, therefore, its expansion is directed as well. Our results also show that human decisions affect the species expansion.

**Key words:** Generalized linear model, invasive species, *Robinia pseudacacia*, spatial distribution, Prekmurje region.

# Izvleček

Robinija *(Robinia pseudacacia* L.) je bila vnešena v Evropo v začetku 17. stoletja in jo imamo še danes za invazivno rastlino  v več evropskih državah, med drugimi tudi v Sloveniji. Raziskava je potekala v Prekmurju, v severovzhodni Sloveniji.

Cilj pričujoče magistrske naloge je bil odkriti razloge vzorec pojavljanja in razprostranjenosti robinije v Prekmurju. To smo poskušali ugotoviti s pomočjo spremenljivk, ki vplivajo na pojavljanje robinije in napovedati prostorsko razporeditev na treh testnih območjih. Testna območja se nahajajo na Goričkem, na Dolinskem in ob reki Muri. Na raziskovanem območju, ki je obsegalo 12 km$^2$ smo v letu 2009 skartirali sestoje robinije. Določili smo potencialne spremenljivke, ki bi po našem mnenju lahko vplivale na razporeditev te vrste v regiji. Le-te so bile: oddaljenost od najbližje ceste, oddaljenost od najbližjega vodotoka, nadmorska višina, raba tal ter lastnosti tal in kvaliteta zemljišča. Analizo smo izvedli s tehniko prostorskega naključnega vzorčenja, z namenom zbrati opažanja na podlagi razmerij med prisotnostjo robinije in možnimi vplivi spremenljivk. Statistično razmerje smo ugotovili s generaliziranim linearnim modelom (GLM). Končni model je bil uporabljen kot predvidevanje za prostorsko razporeditev vrste na testnih območjih.

Robinija se večinoma pojavlja na travnikih in pašnikih. Bližina do cestnega omrežja se zdi, da olajša prisotnost robinije do določene mere. Zrazdaljo od vodotoka se zmanjšuje prisotnost te vrste. Soodvisnosti med nadmorsko višino in prisotnostjo robinije pa nismo zaznali. To dokazuje, da očitno nadmorska višine ne vpliva na razširjenje robinije v regiji. Robinija se širi po naravni poti, vendar da jo kmetovalci tudi namenoma zasajajo, tako je širjenje povezano tudi s človeškimi dejavniki. Naši rezultati tudi dokazujejo, da človekove dejavnosti močno vplivajo na razširjanje vrste.

**Ključne besede:** generalizirani linearni model, invazivna vrsta, *Robinia pseudacacia*, prostorska razporeditev, Prekmurje.

## Abbreviations

ANN = Artificial Neural Network

ANOVA = Analysis of Variance

AUC = Area Under the Curve

BRT = Boosted Regression Tree

CART = Classification and Regression Tree

DEM = Digital Elevation Model

DOF = Digital Ortho Photo

ENFA = Ecological Niche Factor Analysis

FAO = Food and Agriculture Organization of the United Nations

GAM = Generalized Additive Model

GIS = Geographic Information System

GLM = Generalized Linear Model

MARS = Multivariate Adaptive Regression Splines

MAXENT = Maximum Entropy Modelling

ROC = Receiver Operating Characteristic

TPR = True Positive Rate

TSS = True Skill Statistic

WRB = World Reference Base for Soil Resources

# Index

Abstract ........ ii

Izvleček ........ iii

Abbreviations ........ iv

Index ........ v

Index of Figures ........ vii

Index of Tables ........ viii

About the author ........ ix

1. Introduction ........ 1
   - 1.1. Overview of studies done in habitat modelling of invasive species ........ 3
   - 1.2. Research problem ........ 5
   - 1.3. Research hypotheses ........ 8
   - 1.4. Thesis Outline ........ 8

2. Theoretical framework ........ 9
   - 2.1. Description of *R. pseudacacia* ........ 9
   - 2.2. Ecological niche of *R. pseudacacia* ........ 11
   - 2.3. Ecological niche modelling of invasive species ........ 12
   - 2.4. Ecological niche modelling approach ........ 14

3. Materials and Methods ........ 16
   - 3.1. Description of study area ........ 16
   - 3.2. Description of vegetation in the region ........ 20
   - 3.3. Description of datasets used and data pre-processing ........ 22
   - 3.4. Experimental design ........ 29

4. Results ........ 35
   - 4.1. Modelling approach and prediction ........ 35
   - 4.2. Evaluation of the potential habitat model ........ 46

5. Discussion ........................................ 52
5.1. Are land use and other environmental variables co-acting in determining the current spatial distribution of *R. pseudacacia* in Prekmurje region? ........................................ 52
5.2. Can we predict a potential habitat of *R. pseudacacia* for other areas based on the current habitat of the species in a small fragment of the investigation area? ........................................ 54
5.3. Modelling approach ........................................ 56
6. Conclusions ........................................ 57
6.1. Future work ........................................ 57
7. Summary ........................................ 59
8. References ........................................ 61
Acknowledgements ........................................ 74

# Index of Figures

Figure 1- Typical *R. pseudacacia* stand in the training site. ..................................................... 10
Figure 2- *R. pseudacacia* suckering. ........................................................................................ 10
Figure 3- *R. pseudacacia* flowers............................................................................................ 10
Figure 4- Scheme of the life history of R. pseudacacia (Source: Sakio n.d.)............................. 11
Figure 5- Location of the study area. (The orange rectangle represents the training site, the yellow square the testing site in the hilly area, the brown square the testing site in the lowland area and the green square the testing site in the floodplains.) ...................................... 18
Figure 6- Graphical summary of Land use types in the *training dataset.* .................................. 19
Figure 7-Graphical summary of Soil types in the *training dataset.* ........................................... 19
Figure 8- Schematic view of the forest vegetation of Pomurje region (source: Čarni et al. 2008). ........................................................................................................................ 21
Figure 9- *R. pseudacacia* stands (red colour polygons) in the training site mapped on field survey. ........................................................................................................................ 23
Figure 10- Sampling design used to build the model (in blue colour the points within *R. pseudacacia* stands and beige colour the points outside *R. pseudacacia* stands). ...................... 28
Figure 11 – Scheme representing the relationships among the data, the model performed and the prediction done in order to obtain a potential habitat of *R. pseudacacia* for the testing sites (modified from Fukasawa et al. 2009). ............................................................... 30
Figure 12- Relationship curve between distance from water and the target species fitted by GLM. .......................................................................................................................... 37
Figure 13- Relationship curve between distance from road and the target species fitted by GLM. .......................................................................................................................... 38
Figure 14- Probability of occurrence for R. pseudacacia in the training site. ............................ 39
Figure 15- Occurrence probabilities of *R. pseudacacia* for the lowland testing site as predicted by GLM using environmental predictors. ................................................................. 41
Figure 16- Occurrence probabilities of *R. pseudacacia* for the Mura floodplains testing site as predicted by GLM using environmental predictors. .............................................................. 43
Figure 17- Occurrence probabilities of *R. pseudacacia* for the Goričko testing site as predicted by GLM using environmental predictors. ................................................................. 45
Figure 18- Plot of the True Positive Rate result from the *training site.* ..................................... 46
Figure 19- Plot of the True Positive Rate results for the three *testing sites.* .............................. 47

# Index of Tables

Table 1: Summary of the predictor variables used in the logistic model. .................................. 27
Table 2: Result from the variable selection procedure. Where Df represents the degrees of fredom; Deviance is the likelihood ratio chi-square value; and the significance p-value in the is based on the change in Deviance and the change in Df produced by adding that term into the model. Asterisks refer to significance level: '***' - p< 0.001, '**' - p<0.01, '*' - p< 0.05, '.' - p<0.1. .......................................................................................................................... 35
Table 3: Regression coefficients of the GLM. The linear and quadratic terms of each predictor variable were selected when they produced a significance at a 0.05 level. Asterisks refer to significance level: '***' - p< 0.001, '**' - p<0.01, '*' - p< 0.05, '.' - p<0.1................................36
Table 4: Accuracy measurement result for the training site...................................................... 48
Table 5: Accuracy measurement result for the lowland testing site........................................... 48
Table 6: Accuracy measurement result for the Mura floodplains testing site. ............................ 49
Table 7: Accuracy measurement result for the Goričko testing site. ......................................... 49
Table 8: Pairwise comparison matrix for the presence /absence [1/0] of *R. pseudacacia* observed and predicted for the *training site* according to TSS. ................................................. 50
Table 9: Pairwise comparison matrix for the presence /absence [1/0] of *R. pseudacacia* observed and predicted for the lowland *testing site* according to TSS. ..................................... 50
Table 10: Pairwise comparison matrix for the presence /absence [1/0] of *R. pseudacacia* observed and predicted for the Mura floodplain *testing site* according to TSS. ......................... 50
Table 11: Pairwise comparison matrix for the presence /absence [1/0] of *R. pseudacacia* observed and predicted for the Goričko *testing site* according to TSS. ..................................... 51
Table 12: Comparison of TPR curves by paired-sample ANOVA with Tukey post test............ 51
Asterisks refer to significance level: '***' - p< 0.001, '**' - p<0.01, '*' - p< 0.05, '.' - p<0.1. .... 51
Table 13 : Comparison between *R. pseudacacia* stand divided by the limit of soil type classification ........................................................................................................... 55

## About the author

In July 2007 I graduated with a university degree in Environment and Natural Resources Engineering at University of Trás-os-Montes and Alto Douro, in Vila Real, Portugal. Thereafter, I had an opportunity to receive a mobility grant under Erasmus Programme to study abroad at Comenius University in Bratislava, Slovakia for one school year. At Comenius University I attended subjects lectured by great European researchers as, for example, Prof. Dr. Maria Kozová, deputy head of the Department of Landscape Ecology. During my academic studies I was engaged in several courses related to environmental protection, eco-sociology and socioeconomic impacts on environment. Since then, these topics became of particular interest and have motivated me to continue with further education and research in this field.

After obtaining my degree, I was awarded a grant under the Leonardo da Vinci Programme, which allowed me to do training at the Centre for Scientific Research of the Slovenian Academy of Sciences and Arts (ZRC-SAZU), in Ljubljana, Slovenia. Based on field investigation and landscape data, the prediction map of the most invasive plant species *Fallopia japonica* around Ljubljana community was developed, using Geographic Information Systems. The results were presented at the 32$^{nd}$ Symposium of the Eastern Alpine and Dinaric Society for Vegetation Ecology at Pörtschach, Austria in 2009.

After completing this training period, I was employed at the University of Nova Gorica, Slovenia where I worked for the Central European project “Transnational Ecological Networks in Central Europe-TransEcoNet” from January 2009 until January 2011. At the same time, and with an eye on my research career, I felt the need to take a step forward with my education. In November 2008 I engaged in the post-graduate studies, program of Environmental Sciences at School for Environmental Sciences, University of Nova Gorica.

As the topic of invasive species aroused great interest for me during my training at ZRC-SAZU I decided to continue my education in this field with hope to better understand the invasion processes in order to provide solutions for this environmental issue. I started to search for invasive species data in Slovenia and I found that *Robinia pseudacacia* was acting as invasive in many European countries as well as in Slovenia. Due to the availability of a vegetation map for the Prekmurje region with *Robinia pseudacacia* presence the research undertaken was conducted in this region.

During my research work I received a Staff Mobility Grant from the Erasmus Life Learning Programme, which allowed me to work at the Department of Plant Taxonomy and Ecology at Eötvös Loránd University in Budapest, Hungary. Under supervision of dr. Imelda Somodi I started to study the effect of landscape variables in the spatial distribution of *Robinia pseudacacia* in the NE Slovenia, which was the basis of this master thesis.

# 1. Introduction

Recently attention is being directed towards the issue of biological invasions. The migration of species has always existed, however the rate of human-assisted introductions of new species is currently significantly higher than previously (Ricciardi 2007; Vitousek et al. 1997; Lodge 1993; Noble 1989; Pimentel 2005).

Invasive plants, are naturalized plants that produce reproductive offspring, often in large numbers, at considerable distances from parent plants and have a potential to spread over a considerable area (Richardson et al. 2000).

Globalization has greatly increased the speed in the spread of invasive species through economics and trades (Pimentel 2005; Essl et al. 2011); climate change and changes in atmospheric conditions; tourism; transport of goods across borders; land use changes; conflict and reconstruction; regulatory regimes; biological control programmes; public health and environmental concerns (FAO). Due to socio-cultural; economic and environmental harms worldwide biological invasions have been increasingly recognized as important element of global change (Dukes & Mooney 1999; Vitousek et al. 1996).

Global climate change has many environmental consequences, including changes in species distributions and in their abundance within existing distributions. This is due to direct physiological impacts on individual species and changes in abiotic factors, reproduction and recruitment opportunities, and interspecific interactions (FAO).

Many studies (Dukes & Mooney 1999; Thuiller et al. 2005; Walter et al. 2005; Kleinbauer et al. 2010) have proved that climate change might drive the expansion of invasive species as it may produce favourable conditions for its spread.

Once dominant species in native areas are no longer adapted to the environmental changes of their habitat, it is likely that non-native species will displace them thus drastically changing successional patterns, ecosystem function and resource distribution (FAO). These changes may also alter production patterns and trade in agricultural and forestry commodities by species being grown more competitively in higher latitudes and altitudes. Since invasive species establish more easily in habitats disturbed by human and other factors, such changes can provide more opportunities for them to invade (FAO).

Biological invasions interact with land-use change in several ways; however this interaction is not a one-way street (Vitousek et al. 1997). Introduced plants and animals can alter the disturbance regime of sites they invade, for example the introduced fire-promoting grasses have invaded many arid or semi-arid ecosystems, and so have increased the frequency, size and/or intensity of fires (D'Antonio & Vitousek 1992). Also the fragmentation of wild habitats resulting from agricultural or urban development has also affected the spread of introduced species (Vitousek et al. 1997).

According to the FAO conflict and civil unrest can contribute to the introduction and spread of invasive species as a result of various issues for example, the displacement of large numbers of people and their belongings can be a dispersal mechanism for, or the source of, invasive species; the lack of inspections and border controls and the increased unregulated movement of military personnel and refugees; increased smuggling; military transport, equipment and supplies, often covered with dirt or mud from the field, can introduce invasive species into new environments; foreign food aid which may be contaminated with pests and diseases; emergency relief, reconstruction efforts, and humanitarian assistance after wars and disasters.

A country that lacks regulatory regimes, including resources for prevention and enforcement measures regarding risks, are more vulnerable to invasions (FAO).

With millions of tourists crossing international borders every year, the opportunities for the introduction and spread of invasive species is increasing. Travellers can intentionally transport living plant and animal species that can become invasive (FAO).

The intentional importation and release of insects, snails, plant pathogens and nematodes for biological control of pests can be another source of invasive species. Such species can escape into other unwanted areas and become pests themselves (FAO).

Concerns about the effects of pesticides on the environment or/and human health can also promote the spread of invasive species by allowing such species to spread unchecked (FAO; Gurevitch et al. 2006).

Phenomenon of invasion has serious consequences, both for humans and for the preservation of native species and natural communities (Lodge 1993; Vitousek et al. 1997).

Still non-native species must accomplish three steps in order to become invasive. According to Groves (1986) these three stages are introduction, colonization and naturalization. They must find a niche to become established (introduction), grow with necessary resources and survive competition and predation (colonization) and finally spread and reproduce to have subsequent generations (naturalization) (Forman 1995).

Invasive plant species may have the ability to grow rapidly into a new site due to several reasons (Gurevitch et al. 2006). Natural enemies, as competitors and pathogens, which are present and control them in their native range are absent in the new range (Gurevitch et al. 2006; Forman 1995; Pimentel 2005). Human disturbances may reduce the competition with other plants, altering resource availability and allowing the establishment and multiplication of invaders (Gurevitch et al. 2006; Pimentel 2005). The existence of free niches in the new communities in which an invader may be able to proliferate is another reason (Gurevitch et al. 2006; Hierro et al 2005). For example a gap in a forest canopy may allow the introduction of new species into a forest stand. Finally, the invaders may alter the characteristics of an ecosystem in order to favour their own increase (Gurevitch et al. 2006).

Invasive plant species may have negative effects on the retention and restoration of native species and communities; some studies have proved that introduced species have modified the native vegetation from the new environment (Allan 1961; Walker & Vitousek 1991). Some ecologists also believe that plant invasions contribute substantially to plant species extinctions (Wilcove et al. 1998; Vitousek 1994; Mooney & Drake 1989), while others defend that this relation is poorly documented (Forman 1995; Sax et al. 2002; Gurevitch & Padilla 2004). Invasive species can decrease native biodiversity and disrupt ecosystem processes by altering the nitrogen cycle, fire regimes, and other characteristics of ecosystems (Higgins et al. 1999; Vitousek et al. 1996; D'Antonio & Mack 2001). Biological invasions can lead to enormous economic costs, including decreases in timber growth rates and forest values if forest ecosystems are invaded (Gurevitch et al. 2006; Pimentel 2005).

In invasive species control, the data and mathematical analysis of the data are important for decision-making management, and statistical methods play a crucial role.

## 1.1. Overview of studies done in habitat modelling of invasive species

The potential habitat of invasive species is usually predicted by the use of "habitat suitability model", also known as "ecological niche model", which explains the occurrence of the species as a function of environmental factors (Higgins et al. 1999;

Fukasawa et al. 2009). This approach is based on the observed correlation between environmental factors and the distribution of the species (Fukasawa et al. 2009).

Several studies have been focused in habitat modelling of invasive species; these studies have suggested that some habitat attributes may predict future invasions, but the degree of predictability varies according to the type of statistics and the variables used (Marcot 2006). Some have used geographic information system (GIS) techniques (Johnson et al. 2001; Gillham et al. 2004; Sebert-Cuvillier et al. 2008), statistical methods such as Ecological Niche Factor Analysis (ENFA) (Strubb & Matthysen 2009), regression approaches (e.g. GLM and GAM) (Higgins et al. 1999; Guisan et al. 2007; Mingyang et al. 2008; Siapatis et al. 2008; Dullinger et al. 2009; Fukasawa et al. 2009; Hortal et al. 2010; Guisan & Zimmermann 2000), maximum entropy method (Maxent) (Guisan et al. 2007; Mingyang et al. 2008; Holcombe et al. 2010; Hortal et al. 2010; Trethowan et al. 2011), boosted regression Trees (BRT) (Guisan et al. 2007), classification and regression tree (CART) (Mingyang et al. 2008), multivariate adaptive regression splines (MARS) (Guisan et al. 2007), artificial neural networks (ANN) (Lippitt et al. 2008), BIOCLIM (Guisan et al. 2007), genetic algorithms (GARP) (Peterson et al. 2003; Guisan et al. 2007; Mingyang et al. 2008; Dudei & Stigall 2010), among others.

Some of these studies have as a source of information presence-only data to fit models (e.g. Guisan et al. 2007; Peterson et al. 2003), this kind of data is commonly compiled from collections in museums, herbaria and national biological data centres, and are often recorded without a sampling strategy (Guisan et al. 2007). This type of data normally settles pseudo-absences data, this data is usually sampled randomly from among points across the study area where the species has not been detected (Peterson 2003). While others studies have used presence-absence data (e.g. Higgins et al. 1999; Hortal et al. 2010), the valid absences can be very important for the model building. The species data source can be also used as abundance observations (Guisan & Thuiller 2005).

The common approach to all these techniques is that they include species occurrence data combined with environmental factors to develop a potential distribution of the species which can be represented, with use of GIS, as a predictive map for species distribution.

In general, the methods used in these studies are promising for indicating the potential distribution of species’ invasions; however, some limitations should be mentioned.

The main problem of these studies is that the models only apply to the study area, normally local or regional, and generally these models are not applicable to areas outside the study areas (e.g. Fukasawa et. al. 2009). This can be due to the limited availability of occurrence records; such as when data are available for limited number of species (Peterson 2003). Broad-scale information networks are still lagging behind.

The source of information on species' distribution can be also an issue, the importance of presence-only records as a data source demands attention to determine the best methods for modelling them, requiring the use of specific techniques.

In presence-absence models, two kinds of predictions errors are possible. These are commission errors (arise from predicting a species where it doesn't occur) and omission errors (failing to predict a species where it does occur) (Guisan & Thuiller 2005).

The properties of occurrence data might affect the model performance (Guisan et al. 2007).

## 1.2. Research problem

Slovenia is among the European countries with most preserved nature and the highest biodiversity. This is reflected by the size of the Slovenian territory that is protected, 52% lies within ecologically important areas and approximately 35 % within Natura 2000 sites. However Slovenia suffers also many pressures, one of them are the biological invasions. Over the past decade, the share of invasive species has been increasing. So far, between 30 to 60 plant species in Slovenia are considered as invasive, and *R. pseudacacia* is among the worst invasive species (Jogan 2009). They are mainly confined to human settlements and areas where they are easily transported, like rivers and roads (Jogan 2001). The large densities of invasive plant species are found in the southwest (around the city of Koper), west (around the city of Nova Gorica), central Slovenia (around Ljubljana and vicinity), southeast (around the city of Novo Mesto) and north east (around the city of Maribor), which are the largest cities of Slovenia (Jogan 2001). In the North West the largest stripe of invasive plant species is along the Mura River (Jogan 2001). The places with the lowest density of invasive species are the mountainous areas, especially the Alps and the Dinaric Mountains (ARSO).

The species chosen for this study was *Robinia pseudacacia* L., this species was selected due to its abundance in the study region and it has been present in the country for over 100 years (Novice 1858). The species is also known by its common names, among others, as black locust, locust tree and false acacia.

*R. pseudacacia*, a pioneer tree native to North America (Morimoto et al. 2009; Kleinbauer et al. 2010), was introduced into Europe at the beginning of the 17$^{th}$ century by Jean Robin, a major botanist at the Jardin des Plantes in Paris, as ornamental tree and is now being considered to be an invasive species in many European countries (Michener 1988; Kowarik 2003; Walter et al. 2005; Bartha et al. 2008; Kleinbauer et al. 2010) including Slovenia (Wraber 1964; Jogan 2000; Rudolf & Brus 2006).

Although *R. pseudacacia* is considered as invasive plant (Pyšek et al. 1998; Jogan 2000; Morimoto et al. 2009; Kleinbauer et al. 2010), the distribution trend in Europe is increasing and it is the most widely planted tree from America (DAISIE 2006). It is still being planted due to its economic importance, such as timber (Huntley 1990; Rédei et al. 2001; Rudolf & Brus 2006), honey production (Huntley 1990; Wieseler 1998; Rédei et al. 2001; DAISIE 2006; Rudolf & Brus 2006), soil erosion control (Ball 1968; Wieseler 1998; DAISIE 2006), ornamental use (Ball 1968; Huntley 1990; Rédei et al. 2001; Rudolf & Brus 2006), and homeotherapy uses due to essences derived from its flowers (Bartha et al. 2008).

The advantages (e.g. vineyards poles) of planting this species in Slovenia were already mentioned in 1858, as well the differences between young and old stands (Novice 1858). This allows us to infer that the species was introduced to Slovenia in the early 19$^{th}$ century. It has subsequently spread via plantations and natural propagation (Jogan 2000; Rudolf & Brus 2006). At the moment *R. pseudacacia* is the most common introduced tree species in Slovenian forests (Rudolf & Brus 2006).

The species spread has impacts on ecosystems, human health, society and economy. Once introduced in an area, the tree rapidly expands (Wraber 1964; Nascimbene & Marini 2010) preventing other trees from becoming established (Hruška 1991). After cutting the other trees or the shrub layer, *R. pseudacacia* invades the area and the local tree species cannot compete (Hruška 1991). The *R. pseudacacia* stands are mostly monodominant; the species has allelopathic effect inhibiting the growth and development of native plants (Wraber 1964; Bartha et al. 2008). Kowarik (1995) have showed that *R. pseudacacia* changes dry grasslands in multiple ways, decreasing the species numbers, and changing the spatial structure and microclimatic situation by the formation of a tree-layer. With its nitrogen fixation roots it enriches the soil promoting the growth of fast-growing plants which many of them are weeds (Veenvliet et al. 2009).

The large and fragrant blossoms of black locust compete with native plants for pollinating bees (Wieseler 1998). Severely altering the nutrient cycles and increasing the productivity of nutrient-poor habitats (Rice et al. 2004; Walter et al. 2005). The tree makes large roots near the surface, sometimes buckling sidewalks or interfering with mowing. Once established it is very difficult to eradicate.

We faced some difficulties to gather information about *R. pseucacacia*, especially studies focusing on the distribution of the species. Few can be found about how to manage and improve *R. pseudacacia* forests, which enforce the advantages of the species (Redei et al. 2001). Although *R. pseudacacia* is an invasive plant rapidly expanding with negative impacts on different ecosystems, there are only a few attempts to define the potential habitat that could favour the spread of the species, e.g. Kleinbauer et al. 2010, which aimed to assess the extent of the species into reserve networks under a climate warming and Krízsik & Körmöczi (2000), which studied the spatial spread of *R. pseudacacia* and *Populus alba* clones in sandy habitats of the Great Hungarian Plain.

In this study, we identified the critical influencing factors for the current occurrence pattern of the species in the training site and predicted the potential habitat of *R. pseudacacia* for unsampled sites (hereinafter testing sites). We chose generalized linear model for developing a testable habitat model to represent and predict the distribution of the *R. pseudacacia* in the study region. This model type was selected because it fulfils the requirements to achieve the proposed objectives and its intended use. Existing knowledge regarding the distribution of the study species is poor, thus the selected variables were used according to our knowledge based on the field observations.

The results can be conveyed to management authorities in terms of the degree to which specific habitat variables, and management activities affecting these variables, can be expected to influence the distribution of the species.

## 1.3. Research hypotheses

H1) Land use and environmental variables co-act in determining the current spatial distribution of *R. pseudacacia* in Prekmurje region;

H2) Based on the current habitat of the species in a small fragment of the investigation area we can predict distribution of *R. pseudacacia* for other areas.

## 1.4. Thesis Outline

Regarding the structure this thesis is divided into 8 main sections.

In Section 1 we provide the introductory notes on the topic of the thesis and give an overview of studies done on the field of habitat modelling of invasive species and the techniques used. A description of the thesis theme, the purpose of this work and the proposed hypotheses are also present in this section.

In Section 2 we provide a theoretical framework for this work, where we describe the study species and its ecological niche; we define ecological niche modelling, its importance to predict the potential habitat of invasive species and its technical aspects; finally the background of the technique used to model the potential habitat of *R. pseudacacia.*

In Section 3 we describe the materials and methods used in this work. We describe the study area with focus on the training and testing sites; the vegetation present in the study region; the used datasets and the followed methodology.

In Section 4 we give the results of the potential habitat modelling.

The Section 5 discusses the obtained results, the issues that appeared during the work and its possible solutions.

Conclusions are drawn in the Section 6, where we also give some ideas for future work.

Finally, in Section 7 we present a short summary of the work done.

# 2. Theoretical framework

## 2.1. Description of *R. pseudacacia*

*R. pseudacacia* is a deciduous tree that belongs to the *Fabaceae* family. It has alternate leaves; compound, odd-pinnate with 3-10 pairs, round leaflets per leaf (Ball 1968). At the base of most leaves a pair of long, stipular spines are found (Ball 1968; Swearingen 2009).

The species produce inflorescences which are a pendant large raceme, 10-20cm, of white (see Figure 3) highly scented flowers, 15-20 mm, and bare glabrous legumes, 5-10 cm long, which remains attached until splitting open in winter (Ball 1968; DAISIE 2006; Bartha et al. 2008). Some legumes open on the trees, others are dispersed by wind (DAISIE 2006; Masaka & Yamada 2009). Flowering occurs in the spring, drooping clusters in May and June (Wieseler 1998).

Although black locust produces abundant seeds the germination of these seeds is low. The species mainly reproduces by root suckering (from established root systems; see Figure 2) and stump sprouting to form a connected root system (Wieseler 1998). Root suckers emerge from the horizontal roots either when cut by humans or under natural conditions. Due to its strong roots, it strengthens the sandy soils and embankments (Zavod za Gozdove Slovenije 2009).

As nitrogen fixing species, living in symbiosis with nodule bacteria (Walter et al. 2005), black locust is an early successional plant which achieves dominance on open sites where nitrogen is limited to other species.

In forests stands it can grow up to 30-35 m in height (Wieseler 1998) and in favourable conditions, reaches the age of 200 years (Zavod za Gozdove Slovenije 2009). In juvenile trees the bark is smooth and green with suberous lenticels, while in older trees the bark is very thick and grey-brown, yellowish in the cracks (Wieseler 1998; see Figure 1).

Figure 4 shows a scheme of the life history of *R. pseudacacia*.

**Figure 1-** Typical *R. pseudacacia* stand in the training site.

**Figure 2-** *R. pseudacacia* suckering.

**Figure 3-** *R. pseudacacia* flowers.

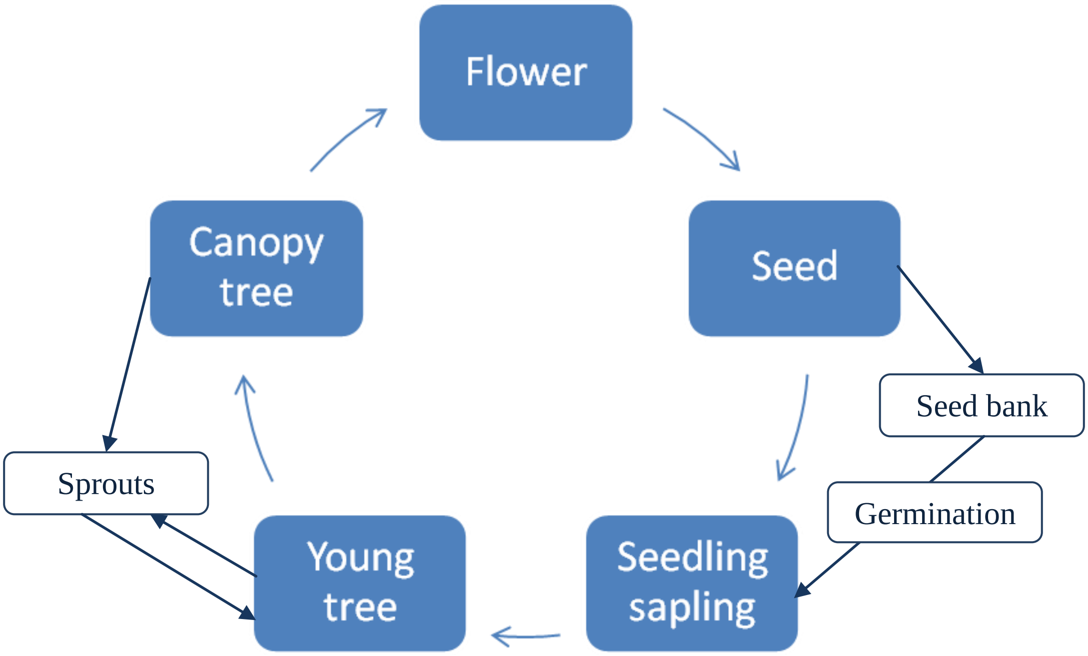


**Figure 4-** Scheme of the life history of *R. pseudacacia* (Source: Sakio n.d.).

## 2.2. Ecological niche of *R. pseudacacia*

In its native habitat the tree grows on moist, limestone-derived soils, in upland oak-hickory forests. In invaded ranges the tree occupies diverse habitats as surface running waters; grassland and tall forb habitats; lines of trees, small anthropogenic woodlands, recently felled woodland, early-stage woodland and coppice and transport networks and other constructed hard-surfaced areas (DAISIE 2006). In central and western Europe it particularly invades nutrient poor dry and semi-dry environments (Kleinbauer et al. 2010). It can be also found in disturbed areas such as old fields, degraded woods, and roadsides (Wieseler 1998). In the Great Hungarian Plain it is successful in the occupation of sandy habitats, disturbed habitats and it more aggressive after a forest fire (Krízsik & Körmöczi 2000).

As nitrogen fixing species, is an early successional plant which achieves dominance on open sites where nitrogen is limited to other species and it rapidly colonizes acidic or polluted soils (DAISIE 2006). It can also grow on salty soils (Torelli 2002).

Although it grows in a wide range of sites it cannot grow in the shade, only in the first 6-8 years can it survive with some shade. It grows well in full sun and well-drained soils, and it is drought tolerant. It is not well adapted to high ground water levels or stagnant water (Rudolf & Brus 2006).

In Slovenia the species is found on warmer and lowland areas (Jogan 2000; Rudolf & Brus 2006). It invades forest edges, forest clearings, non-dense forests (Wraber 1964) and abandoned surfaces (Rudolf & Brus 2006).

This species has large ecological niche (Kleinbauer et al. 2010), i.e. is “generalist”, grows rapidly and exhibits clonal spread what increases its distribution into the surrounding landscapes. Thus is important to understand the patterns of distribution of the species to prevent future invasions.

It is also important to mention that the study species presents an r reproductive strategy, typical from pioneer plants, it produces large amount of seeds, has also ability and dynamics of clonal growth,  sprouts emerge earlier than native species (Krízsik & Körmöczi 2000).

## 2.3. Ecological niche modelling of invasive species

The term of “ecological niche” stands for the position of a species within an ecosystem, describing both the range of conditions necessary for the occurrence of the species, and its ecological role in the ecosystem (Polechová & Storch 2008).

There are many approaches to the ecological niche and sometimes the ecological niche is also described as species’ habitat requirements (Polechová & Storch 2008). Therefore in ecological niche modelling the terms potential habitat modelling, potential distribution modelling, spatial distribution modelling, among others, are often used.

It is important to distinguish between “fundamental” and “realised ecological niche”, particularly with regard to the methodologies used. The “fundamental ecological niche” of a species is the area that comprises the environmental (abiotic and biotic) conditions within which a species can survive and grow (Soberón & Peterson 2005); the “realised ecological niche” is normally a smaller area than the former, in which a species is excluded from parts of its *fundamental* niche due to competition and other biotic interactions (Guisan & Zimmermann 2000; Soberón & Peterson 2005).

Two approaches have been developed in ecological niche modelling: (1) "Mechanistic approach" to the niche modelling as referred by Guisan & Zimmermann (2000). This approach consists on the direct measurement or physical modelling of the species responses to the physical parameters, and inferring from them fitness values of different combinations of physical variables (Pearson & Dawson 2003; Soberón & Peterson 2005). Then the geographic regions of positive fitness can be displayed with the use of GIS technology (Porter et al. 2000). (2) "Correlative approach" to the ecological niche modelling assumes that the niches may be reconstructed by relating data on species' occurrences with combinations of environmental variables (Pearson & Dawson 2003; Soberón & Peterson 2005). This is done by quantifying the relationship between occurrence of the species and the set of environmental factors by determining the correlation between them (Guisan & Thuiller 2005). In essence, this approach extrapolates from the associations between species' occurrences and environmental variables to identify areas of predicted presence on the map. These areas are ecologically similar to those where the species is known to occur (Soberón & Peterson 2005).

The mechanistic approach, being based on direct measures of physiological variables, ignores the biotic interactions. While, the correlative approach is based on observations that already include the effects of biotic interactions on the distributions of species (Soberón & Peterson 2005). Thus these two approaches estimate quite different phenomena.

The ecological niche models that are calibrated using species' occurrence data are based on the "realised ecological niche" rather than the "fundamental niche" (Pearson & Dawson 2003; Trethowan et al. 2011). The models that aim to identify the "fundamental niche" look for a more physiologically based mechanistic approach (Pearson & Dawson 2003). It is important to mention that predicting future species distributions based on physiology determines fundamental niches that are unlikely to be as accurate as those based on correlations between the species distribution and the current realized niche (Pearson & Dawson 2003).

The result of the ecological niche modelling is a potential habitat map of the invasive species in question. The mapping of invasive species distribution allows for management actions to be taken in areas that are suitable for the invasive species (Thuiller et al. 2005) and is fundamental for early warning systems to develop preventive policies (Zhu et al. 2007).

## 2.4. Ecological niche modelling approach

The "generalized regression" modelling methods is a broad family of methods that are the most often applied approach to species modelling and included Generalized linear models (GLM), Generalized additive models (GAM), vector GLM and GAM (VGLM/VGAM), multiple additive regression splines (MARS) and generalized linear and additive mixed models (GLMM/GAMM) (Guisan et al. 2006).

Regression analysis is used for explaining or modelling the relationship between a single variable Y, called response, output or dependent variable and one or more predictor or explanatory variables $X_1, \ldots, X_p$. It is called simple univariate regression when $p=1$, and multiple regression or multivariate regression when $p>1$ (Faraway 2004; Hastie et al. 2009). The explanatory variables can be continuous, discrete or categorical but the response must be continuous (Faraway 2004). Regression analyses include several objectives such as prediction of future observations; assessment of the effect of explanatory variables and response or relationship between them and a general description of data structure (Faraway 2004).

There are three main reasons for using regression techniques: they have wide applicability; they can be most straightforward to implement; and many complex statistical procedures can be better appreciated once regression methods are understood (Kleinbaum et al. 2008).

Generalized linear models (GLM) have the same basic model formulation as linear models, except that they can model data with various statistical distribution properties by introducing a link function between the response and the explanation. The model fitting in generalized linear modelling has to be done iteratively, and distributional results used for inference, are approximate and justified by large sample limiting results, rather than being exact (Wood 2006). The standard linear model cannot handle nonnormal responses such as counts or proportions. This motivated the development of generalized linear models that can represent categorical, binary or other response types (Faraway 2006) and allow for a degree of non-linearity in the model structure (Wood 2006). With GLMs we are able to tackle a wider range of data with different types of response variables (Faraway 2006).

GLM is defined by two components, the response should be a member of the exponential family distribution and the link function describes how the mean of the response and a linear combination of the predictors are related (Faraway 2006).

Logistic regression is a multiple regression with a categorical outcome variable and predictor variables that can be continuous or categorical (Field 2009). When we are trying to predict the relationship of only two categorical outcomes the analysis are known as binary logistic regression, if we want to predict the relationship of more than two categories then we use multinomial logistic regression (Field 2009).

In order to have a valid model in linear regression the observed data should contain a linear relationship (Hastie et al. 2009), when the outcome variable is a categorical, this assumption is violated (Field 2009). One way to solve this problem is to transform the data using the logarithmic transformation; this is a way of expressing a non-linear relationship (multiple linear regression equation) in a linear way (logarithmic terms), called the *logit*, and thus overcome the problem of violating the assumption of linearity (Field 2009).

The value resulting from the equation means the probability that a case belongs to a certain category, therefore this value varies between 0 and 1 (Field 2009). A value close to 0 means that the probability of the event occurs is very unlikely and a value close to 1 means that it is likely to occur (Field 2009). In the regression equation each predictor variable has its own coefficient, which is estimated when we run the analysis. These parameters are estimated by fitting models. The chosen model will be the one that the results in the predictor values are closest to the observed values (Field 2009).

# 3. Materials and Methods

## 3.1. Description of study area

This research was conducted in the Northeast of Slovenia, in the Prekmurje region. This region is considered one of the most important agricultural regions in Slovenia (Gabrovec & Kladnik 1997; Cunder 2009).

The region lies at low altitudes (from 150 to 400 m), reaching the 400 m only in single hill tops in the area of Goričko, in the northern part along the border with Austria and Hungary (Perko & Adamič 1998).

The region is open towards the Pannonian plain and has the most continental climatic features in Slovenia (Ogrin 2009). The average year temperature is 9,6 ºC, the average in January is -1,2 ºC and the average in July is 19,7 ºC (ARSO). The annual amount of rainfall is the lowest in Slovenia (Perko & Adamič 1998; Ogrin 2009) and it is about 805mm (ARSO).

Prekmurje can be divided into three geographical areas, which are as well ecologically dissimilar; the hilly area of Goričko, the floodplains of the Mura River, known as Ravensko, and the lowlands known as Dolinsko.

In Goričko there is no primary forest vegetation, but the forest was settled on former agricultural land. As the relief circumstances (with moderate inclination) were good for agriculture purposes, the forest was slowly removed and those surfaces were converted into agricultural surfaces. When the soil under the agriculture land use was exhausted these areas were naturally or men-induced afforestated. This land use historical cycle was short enough that many locals still remember that the forested surfaces were once agricultural lands. There are still remnants of fields in the forest surfaces (Wraber 1959). The same system of alternating between agriculture and forest land uses was also used in some areas of the lowland (Wraber 1968). The traditional management on Mura River floodplain was the clear cutting and artificial rejuvenation with planting or seeding of forests. Along the rivers Mura and Ledava the biggest forest surfaces, located in areas that were regularly flooded or with high groundwater level, were maintained (Wraber 1968).

Goričko was classified as Natura 2000 site and established as Goričko Landscape Park in 2005. This area is a natural habitat of endangered and rare European flora and fauna species. Its cultural landscape is characterized by self-sufficient farming. The dry grasslands are developed on the tops of the hills and to the exposed sunny slopes of the hills, and the wet meadows are found on the slopes of the hills (Paušič et al. 2011). Ridges and higher surfaces, mainly in central Goričko are covered by sandy clayish soils with quartz sand and gravel. Pliocen gravel in central and eastern Goričko is covered by thin acid soil, which is washed and less fertile. Sandy gravely grounds are usually covered with pines woods. Similarity is also found in northern locations and on steeper slopes (Perko & Adamič 1998).

Holocenic sediments in the lowland area build a yellow and red sandy clay and loam base (Rudolf & Brus 2006). The best agricultural surfaces are settled in this area (Perko & Adamič 1998).

The floodplains of Mura River are also included in Natura 2000 sites. They comprise many meanders, oxbow lakes and flood forested areas (Globevnik & Kaligarič 2005). On sandy and gravely deposits of river Mura brown soil is predominant, which is most appropriate for agriculture (Perko & Adamič 1998), and where are settled the agriculture areas.

In the lowland part of the region a *training site* of 12km$^2$ was chosen to build the model. In order to test the model three *testing sites* were chosen, in the hilly area of Goričko, in the lowland and in the floodplains of Mura River (see Figure 5).

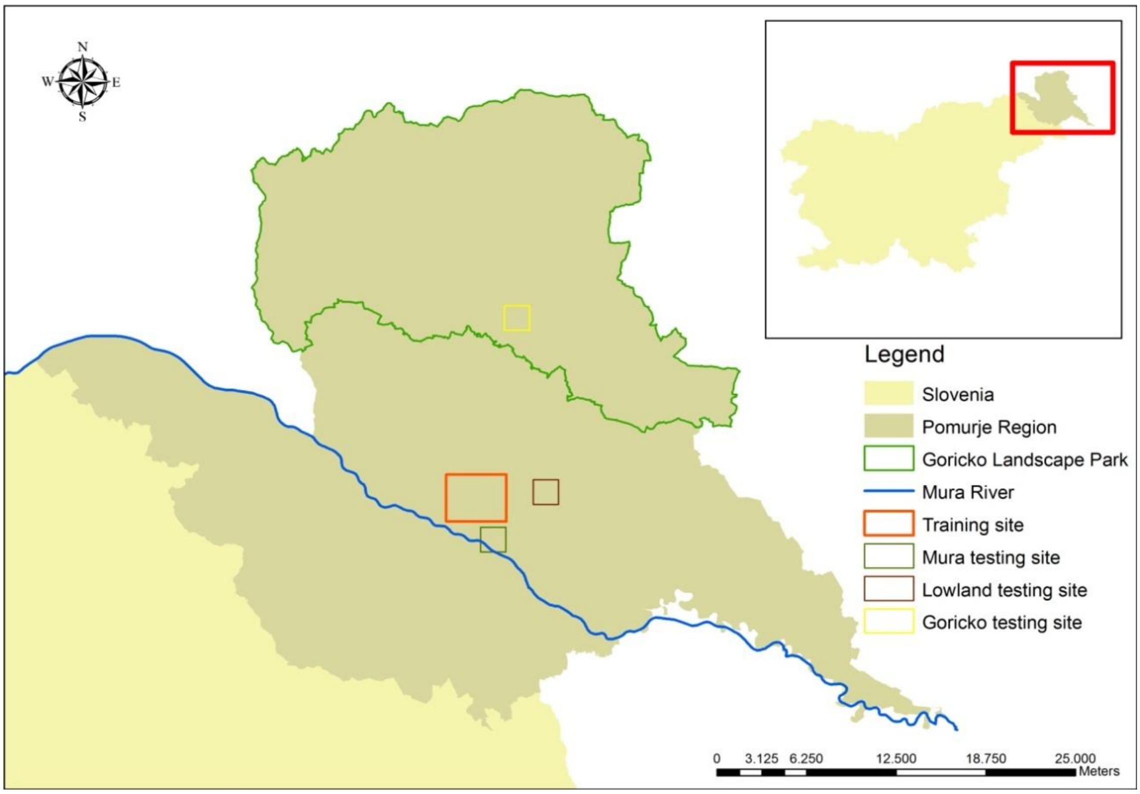


**Figure 5-** Location of the study area. (The orange rectangle represents the training site, the yellow square the testing site in the hilly area, the brown square the testing site in the lowland area and the green square the testing site in the floodplains.)

Approximately half of the *training site* is occupied by meadows, followed by fields. Smaller proportion is occupied by urban areas and similar area by forests. There are fewer surfaces of pastures, and the lower proportion is comprised by orchards. The distribution of land use types in the training area can be seen by the following pie chart (Figure 6).

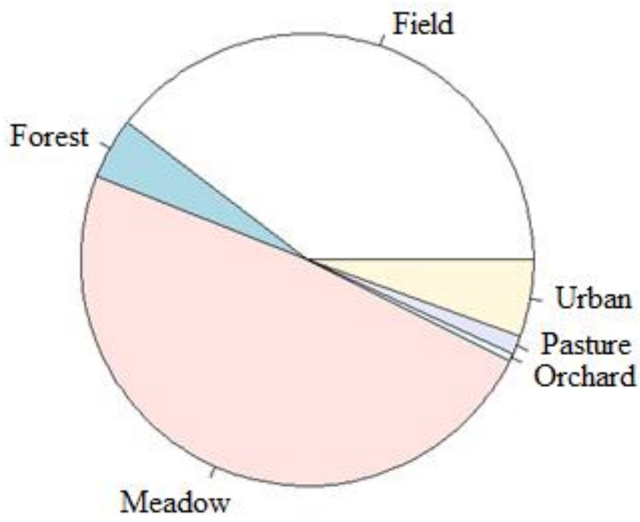


**Figure 6-** Graphical summary of Land use types in the *training dataset.*

Figure 7 shows the ratio distribution of the three soil types present in the *training site*, being the highest part *cambisols*, followed by *fluvisols* and the smaller proportion is due to *Urban soils*.

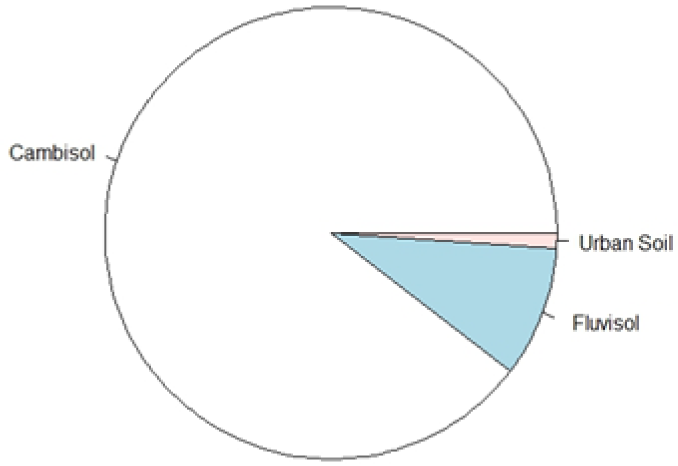


**Figure 7-**Graphical summary of Soil types in the *training dataset.*

The study species can be found in the whole Prekmurje region (Wraber 1964; Rudolf & Brus 2006). Rudolf & Brus (2006) found that the species aggressively regenerates the lowland area, mainly due to the underground water level decrease and strong stand opening, which creates ideal conditions for its development (Rudolf & Brus 2006). The species is also favourable in this area due to the sandy clayish and loam base (Rudolf & Brus 2006).

In Goričko the species is widely distributed, and makes up a very low portion of the growing stocks, except in the wine growing areas, where it is strongly favoured for the production of vineyard poles (Rudolf & Brus 2006). In this area the species invade non-forested areas and overcome forests that are not dense (Wraber 1964). Nevertheless, the acid soil and the hilly topography are not appropriate for the species' occurrence (Rudolf & Brus 2006).

Along the Mura River and its extensive floodplain *R. pseudacacia* shows the largest share of the growing stock (Rudolf & Brus 2006). The species does not coop with high ground water level or stagnant water, however it still grows on these surfaces because they are limestone surfaces which lie on alluvial sands, which are well drained base (Rudolf & Brus 2006).

## 3.2. Description of vegetation in the region

As mentioned above the region under consideration consists of hills, lowland and floodplain along Mura River.

The hills of Goričko are overgrown with beech forests, above all with the acidophilous *Castaneo-Fagetum*, and sporadically with the mesophilous *Vicio oroboidi-Fagetum* (Čarni et al. 2008).

The lowland used to be covered with a mesophilous hornbeam forest of *Pruno padi-Carpinetum* (sporadically with the acidophilous *Vaccinio-Carpinetum*), however nowadays these surfaces have been converted into fields (Čarni et al. 2008). On the marshy surfaces occur alder forests of *Carici elongatae-Alnetum glutinosae* and ash forests of *Pruno padi- Fraxinetum* and *Lonicero caprifolii-Quercetum* (Čarni et al. 2008). In these sites also occur successions of secondary alder forests of *Carici brizoides-Alnetum* (Čarni et al. 2008).

On the banks of Mura River grow floodplain forests, as *Salicetum albae*, *Fraxino-Ulmetum* and *Genisto-Quercetum* (Čarni et al. 2008).

Common in this region, and the object of this work, is the neophytic tree species *Robinia pseudacacia*.

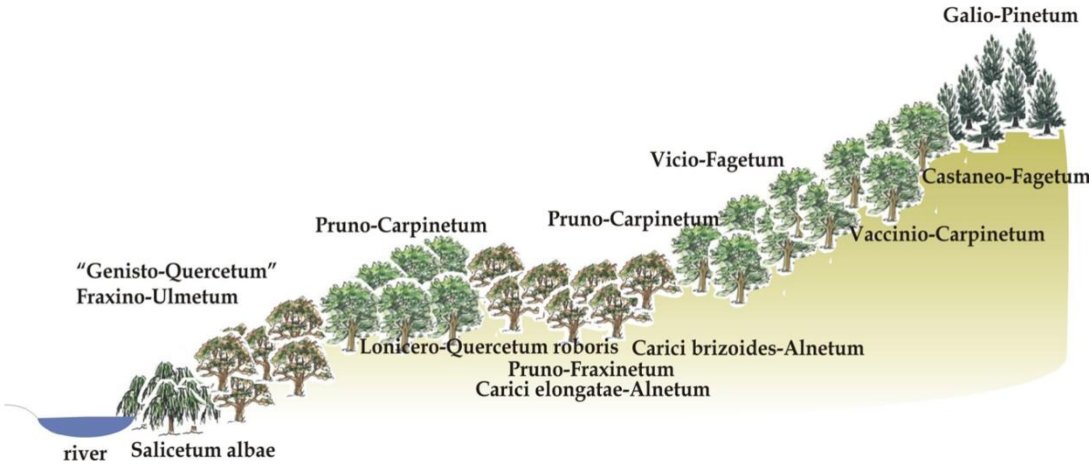


**Figure 8-** Schematic view of the forest vegetation of Pomurje region (source: Čarni et al. 2008).

Wraber described the forest communities in Prekmurje in 1964 and indicated that *R. pseudacacia* is an important species in Querco-roboris Carpinetum community in the lowland area. This result corroborates with the study from Rudolf & Brus (2006), where they document that in the lowland area *R. pseudacacia* can quickly develop into pure stands, especially under improper management, being relatively aggressive in mixed oak-hornbeam forests, which corresponds to the community *Pruno padi-Carpinetum* described by Čarni and colleagues (2008). In the marshy surfaces the species' regeneration is most intensive in the Querco-roboris Carpinetum (Rudolf & Brus 2006) which corresponds to the *Lonicero caprifolii-Quercetum* (Čarni et al. 2008) community, mentioned above.

## 3.3. Description of datasets used and data pre-processing

The initial data analysis consisted of descriptive statistics of the data (including graphic summaries) in order to organise the data into a suitable form for analysis (Faraway 2004). According to Faraway (2004), this is a critical step that should be always performed.

The variables used in the modelling procedure were divided into three types according to the level measurement. The qualitative variables, also referred in literature to as categorical, discrete variables as well as nominal (Hastie et al. 2009), for these variables we obtain the count of the number of time that each type occurs (Faraway 2006). So we considered as categorical the variables Land use and Soil type.

Another variable type according to the measurement was the quantitative, numeric or continuous variable. In this type were included the distance to the nearest water body; distance to the nearest road and elevation.

The third variable type considered was ordinal, in this type there is an ordering between the values, but no metric notion is appropriate (Hastie et al. 2009). Soil quality was considered as ordinal variable since the values of soil quality are ranked and the order of the values is important. Since some soil quality categories were rare, i.e., they were represented by a small number and this was hampering the cross-validation, we merged these categories into the nearest soil quality category that was well represented.

The variables were as well classified according to the orientation, whether the variable was intended to describe or be described by other variables (Kleinbaum et al. 2008). The variable under investigation was to be predicted from other variables so we called it dependent variable (Kleinbaum et al. 2008). All the other variables which we used to describe the dependent variable we called it predictor (Kleinbaum et al. 2008).

Our dependent variable named as “Robinia”, represents the occurrence of the study species, and was coded with 0- 1, where 0 is assigned to the absence of the species and 1 to the presence. The predictors were all the environmental factors that we used to model the potential habitat for *R. pseudacacia*, and can be seen in Table 1.

“NA” has been used as a missing value code in our dataset, which is the missing value code used by R, this was possible because the amount of missing data was small.

### 3.3.1. Occurrence records

In 2009 we surveyed the distribution of *R. pseudacacia* in the training site located at the lowland area that targets the locations where the species was present. The distribution map was generated by manually drawing polygons around the borders of areas in which the species occur. The areas where the species was absent were also mapped, in order to have a full presence-absence species map. DOFs of the study area were used as a backdrop. The drawn polygons were then digitized in ArcGIS software.

*R. pseudacacia* polygons were considered for all the stands where the species was present, independently of its relative density, and as well isolated trees. Cultivated stands of the species were included in the mapping process as *R. pseudacacia* presence. The minimum polygon size mapped was 6m in diameter.

A total of 310 *R. pseudacacia* polygons were mapped in the field, as seen in the Figure 9 with red colour, the remaining areas were mapped as *R. pseudacacia* absence, although they are not mapped in figure bellow.


**Figure 9-** *R. pseudacacia* stands (red colour polygons) in the training site mapped on field survey.

For the testing sites, *R. pseudacacia* was determined from the interpretation of coloured DOFs. The polygons with species' presence were digitized in a geographic information system environment. The areas without definite information regarding the species' occurrence were not mapped since we were not sure about the truly absence. The DOFs used for interpretation date from May 2005 in the scale 1:5000 and were acquired from the Surveying and Mapping Authority of the Republic of Slovenia. The recognition of *R. pseudacacia* by DOF' interpretation was possible because the acquired DOFs were taken during flowering period of the species, so the tree appears with white colouring being distinguishable from other vegetation. The texture of the invaded stands and the experience by the interpreter were also important factors for this step.

### 3.3.2. Predictor variables

The environmental factors considered in our analysis were nearest distance from roads, nearest distance from water bodies, elevation, soil type, soil quality and land use.

These potential explanatory variables were selected because we supposed they represent a range of environmental variables that directly or indirectly influence the performance of our study species (Franklin 1995).

Roads, may act as habitat corridors for plant dispersal (Pauchard & Alaback 2004; Christen & Matlack 2006; Kalwij et al. 2008) which may also lead to invasion of adjacent land and in the case of *R. pseudacacia* has a special influence because the species is frequently planted along roads (Bartha et al. 2008).

Water acts as corridor and source of non-native plant invasions (Stohlgren et al. 1998).

Land use can influence a variety of ecological phenomena, including species movements and spread of disturbance (Turner 1987) being one of the most influential factors determining non-native species distribution (Radosevich et al. 2003; Mosher et al. 2009).

Elevation plays a major role in determining the richness and covers of non-native plant species (Stohlgren et al. 1998) and may physiologically constrain the invasion processes (Pauchard & Alaback 2004).

Soil type reflects the range of soil types that is suitable for the species spread (Huenneke et al. 1990; Stohlgren et al. 1998; Radosevich et al. 2003) and soil quality as the capacity of soil to produce goods such as providing a medium to the plant growth.

Nearest distances from roads and from water bodies were calculated from the road network and water bodies' data layers from the Surveying and Mapping Authority of the Republic of Slovenia.

Land use types were taken from the Land Cadastre Map from the Surveying and Mapping Authority of the Republic of Slovenia.

For this work we classified land use types into six categories: "Field", "Forest", "Meadow", "Orchard", "Pasture" and "Urban", as can be seen in Figure 5. Due to the rarity of some land use types within the *training site* we grouped some land use types; "Vineyards" were included into "Orchard", and all the building areas and areas that are occupied by human activities, e.g. playground, courtyard, etc, were included into "Urban".

Elevation values were masked from the Digital Elevation Model (DEM) to the study site. DEM with 12.5 m x 12.5 m resolution, acquired from the Surveying and Mapping Authority of the Republic of Slovenia, was used. Derivates of elevation, such as aspect and slope, were previously tested and since they did not show any influence in the species distribution, because their range is too small for the study area, were not included as predictor variables.

From the same dataset as land use, information about the soil quality was taken, based on taxation of real estate. The data was acquired from the internet, where the number of each parcel and code of the municipality was introduced to access to the values of taxation (*Boniteta* in Slovenian language) which we consider to be directly related to the soil quality of the parcels.

The soil type information was extracted from the Digital Soil Map of Slovenia, in the scale 1:25000, from the Ministry of Agriculture, Forestry and Food.

According to classification of FAO-WRB reference base (World Reference Base for Soil Resources) we can find three types of soil represented in our *training* and *testing areas*: cambisols, fluvisols and urban soils.

*Cambisols* or brown soils are characterized by slight or moderate weathering of parent material and by absence of appreciable quantities of illuviated clay, organic matter, Al and/or Fe compounds (Duchaufour 1998; WRB 2007). Soils with at least the beginnings of horizon differentiation in the subsoil evident from changes in structure, colour, clay content or carbonate content (Duchaufour 1998; WRB 2007). The mineral materials may have any origin and composition, provided they are favourable for brunification and opposed to leaching (Duchaufour 1998). These soils generally make good agricultural land and are intensively used (WRB 2007).

*Fluvisols* are soils frequently renewed by alluvial material (immature soils on recent materials) (WRB 2007). These soils on recent deposits in valleys have great economic and agricultural importance because they are located on the major floodplain of rivers, where they are rejuvenated often by deposition of new materials during floods, and are characterized by the presence of Phreatic groundwater that circulates and is, therefore, not reducing, and undergoes strong fluctuations in level (Duchaufour 1998; WRB 2007). Profiles with evidence of stratification; weak horizon differentiation but a distinct topsoil horizon may be present (WRB 2007). A reference profile from this soil type is a grey alluvial soil, where the humus is a mull of very variable thickness according to growth of vegetation; the scarcely weathered mineral horizon, without definite structure has very variable texture as sandy, silty, more rarely silty clay (Duchaufour 1998). Generally have non-uniform texture which often shows abrupt variations with depth; a silty alluvial soil frequently rests without transition on a gravelly sand layer, which has disadvantages of forming an obstacle to root penetration and of preventing all capillary movement (Duchaufour 1998). As the groundwater contains dissolved oxygen oxidation-reduction processes are slow, sometimes are manifested in an attenuated manner by appearance of small, rust-coloured mottles; seasonal fluctuations in the water table are large (Duchaufour 1998).

Finally a soil type which is defined by FAO-WRB as *Technosols* and can be also called as *Urban* soils, these are soils with strong human influence containing a significant amount of artefacts or are sealed by technic hard rock (WRB 2007). Artefacts are defined solid or liquid substances created or substantially modified by humans as part of an industrial or artisanal manufacturing process; or brought to the surface by human activity from a depth where they were not influenced by surface processes, with properties substantially different from the environment where they are placed; and have substantially the same properties as when first manufactured, modified or excavated (WRB 2007). This soil type is mostly found in urban and industrial areas, in small areas, although in a complex pattern associated with other groups. Thus, cities, roads, mines, refuse dumps, oil spills, coal fly ash deposits and the like are included in *Technosols* (WRB 2007).

With geographic information system techniques, we extracted all the environmental variables and the presence-absence of *R. pseudacacia* (previously mentioned as occurrence records) at background locations. All the data layers were projected in the Slovenian Coordinate System (D48_Slovenia_TM) with the D_D48 datum.

**Table 1-** Summary of the predictor variables used in the logistic model.

| *Environmental variables* | *Range* |
|---|---|
| Land use | Field, Forest, Meadow, Orchard, Pasture, Urban |
| Soil type | Cambisol, Fluvisol, Urban soil |
| Soil quality | 0-70 |
| Distance to water bodies | 0-1340m |
| Distance to road network | 0-423m |
| Elevation | 176-193m (a.s.l.) |

### 3.3.3. Sampling design

To establish the statistical relationships we used a stratified random sample of three times the number of recorded polygons with presence of the study species. The same number of points was settled for the areas where the species was absent.

The random sampling requires the application of two criteria, each point of the sample has the same probability of being included and the points are chosen independently from each other (Legendre & Legendre 1998; McCune & Grace 2002). This sampling was generated in ArcGIS 9.3, and was used to model the species distribution for the training area.

A total of 1860 random points were generated within *R. pseudacacia* stands, as can be seen in the Figure 10 with blue colour, the same number of points were generated outside *R. pseudacacia* stands, as can be seen in the same figure with beige colour.

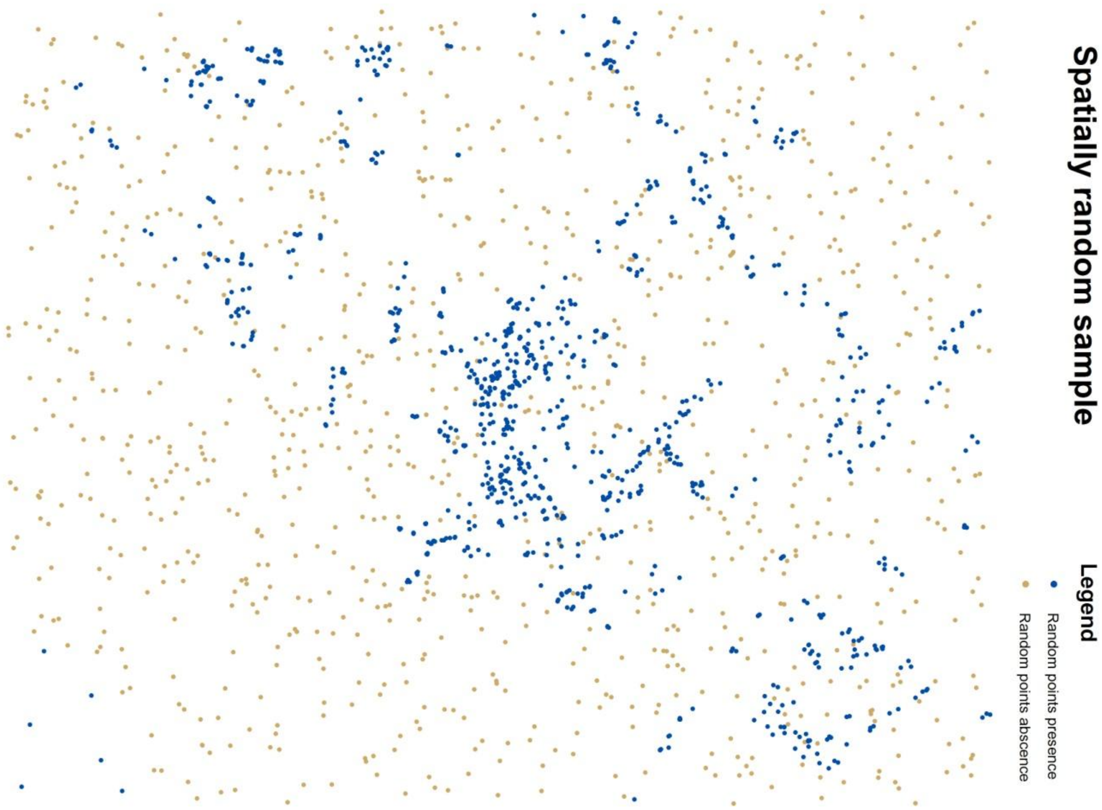


**Figure 10-** Sampling design used to build the model (in blue colour the points within *R. pseudacacia* stands and beige colour the points outside *R. pseudacacia* stands).

For validation a grid of 6m sampling distance, by regular point generation, was used when the model was applied to the training area. Regular sampling means that the sample units (points) are spaced at regular intervals.
Three validation datasets (testing areas) were constructed to assess the accuracy of the model; these datasets sized 4km$^2$ were offset from the training data.

When the predictive model was applied to the testing areas we decided to increase the cell size to 10m, this still provides a sufficient detail in these areas. The final model was then was applied to these grids of points in a predictive mode.

Some points from the testing site on lowland fall into the water bodies, so those points were deleted from the grid, i.e. those were not used in the prediction. This was only possible because they were a relative small number of cases.

## 3.4. Experimental design

We used a GLM to model the potential distribution of *R. pseudacacia*. We modelled presence/ absence of *R. pseudacacia* (dependent variable) using environmental variables (Table 1) as predictors. The linear and quadratic terms were included for all continuous predictors (distances to the nearest water body and the nearest road, and elevation). In order to reduce the quantity of data to be analyzed we random sampled the training and testing sites by a 6m and 10m cell size respectively. Then the information of all the variables (dependent and predictors) was extracted to these sampling points to establish the statistical relationship between species occurrence and environmental factors. Variable selection for GLM was based on comparing candidate models by means of Chi-square tests, which correspond to comparison of alternative linear models by F-test.
The predicted potential habitat of *R. pseudacacia* was done by applying the calibrated habitat suitability model to the three testing sites. The potential habitat was determined as the probability predicted on the basis of the environment at each sampling point.
The accuracy of presence/ absence model predictions was assessed by three approaches: AUC, TPR, and TSS. A Tukey post test was then performed to determine which sites (training and testing sites) differed significantly from one another.
The model was implemented in the R Statistical environment. R is a free software environment for statistical computing and graphics (Faraway 2006). The reasons why we made the statistical model in R are because it is free software, and possesses the power to perform the analyses done in this study. Another advantage of R comparing to other statistical packages is that we can extract all the coefficients individually for subsequent calculations in a convenient way (Faraway 2006).

The result of the potential habitat modelling was projected using GIS as the potential distribution map of *R. pseudacacia* for the three testing sites.

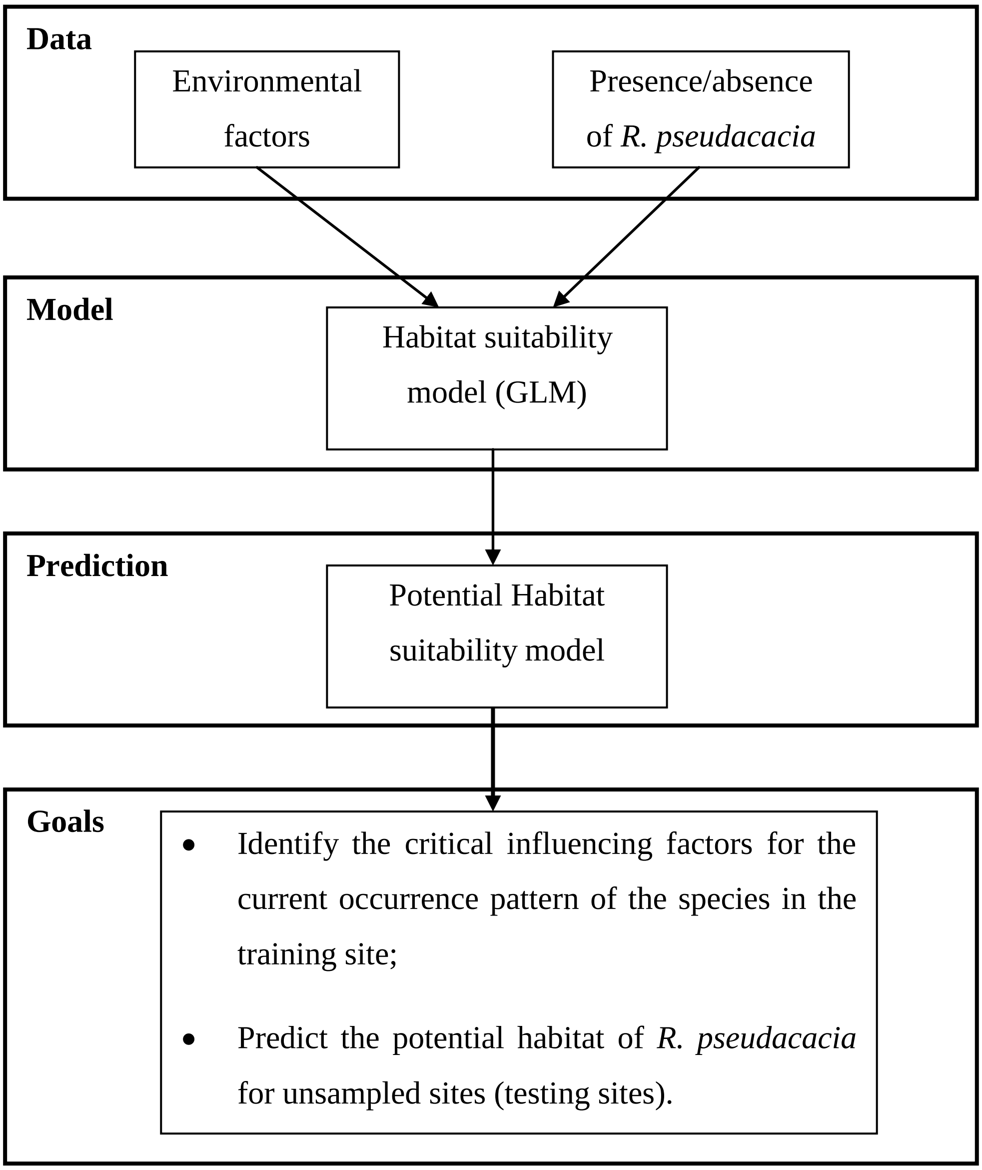


**Figure 11 –** Scheme representing the relationships among the data, the model performed and the prediction done in order to obtain a potential habitat of *R. pseudacacia* for the testing sites (modified from Fukasawa et al. 2009).

### 3.4.1. Modelling approach and prediction

Statistical methods were used to relate the current distribution of the study species with the environmental predictors. To establish this relationship we chose a GLM model.
GLMs constitute a more flexible family of regression models which allows other distributions for the response variable and a non-constant variance functions to be modelled. In this type of model, the combination of predictors is related to the mean of the response variable through a link function. This link function allows (i) the

transformation to linearity and (ii) the predictors to be maintained within the range of coherent values for the response variable. By doing so, GLMs can handle distributions such as the Gaussian, Poisson, Binomial, or Gamma- with the respective link functions, e.g. to identify, logarithm, logit and inverse. These models yield predictions within the limits of observed values (e.g. presence/absence (1/0) and probability values between these extremes) (Guisan & Zimmermann 2000).

The selection of predictors in GLM is usually made automatically by following shrinkage rules. The model is optimized through deviance reduction, which can easily be converted into an estimate. Each model is then generated by multiplying each regression coefficient with its related predictor variable. Once the species' response is derived, its potential distribution can be predicted. The potential distribution maps can be defined as cartographic representation of the probability of occurrence (Guisan & Zimmermann 2000).

The GLM was fitted in R statistical environment (R development team 2008) using the *glm* function in the "stats" package by specifying a binomial distribution and a *logit* link, a canonical link for binomial models, and is the default in R (Wood 2006).

As mentioned above the response variable was represented by the presence/absence of the species and environmental factors were used as predictor variables.

The modelling procedure consisted of three steps, followed by the accuracy.

(1) A correlation matrix among continuous variables was tested to check the independency of the variables. Correlation among variables hampers the model selection, thus leads to wrong error probabilities (Legendre & Fortin 1989). Thus this step aims to check the pairwise correlations between each predictor in the model. The Pearson correlation coefficient was applied to the continuous variables and Cramer's V (statistic measuring of the dependency between two variables) to the categorical variables (Cramér 1946). No two variables with a correlation coefficient higher than 0.8 would have been allowed to enter the model, however we did not verify any correlation among the variables used.

(2) A GLM describing the relationship between predictor variables and the occurrence of *R. pseudacacia* was built.

Linear and quadratic terms were included to the model (Somodi et al. 2010). Second-order polynomials were tested in order to check the non-linear effects of the predictor continuous variables (Dullinger et al. 2009). When the predictor was a first-order

polynomial the probability of *R. pseudacacia* occurring was a sigmoid function, and when the predictor was a second-order polynomial the probability of *R. pseudacacia* occurring was a bell-shaped function (Higgins et al. 1999).
(3) To identify the best subset of variables to include in the model we carried out chi-square model comparison. We used the "drop1" function to perform the selection procedure. All possible models with one variable left out were prepared and a Chi-square test corresponding in design to ANOVA decided the removal of which variable affects the model negatively. We kept only those variables, the lack of which would significantly worsen the model.
The resulting model included the predictor variables with the main effects associated for the occurrence of the species.

When we produced the model based on the sampling data we were wondering if the model would fit the observed data from the training site and whether the model could be generalized to the testing sites. Therefore we used best predictive model to assess the potential habitat of the species at the three testing sites, the potential habitat map was prepared in ArcGIS 9.3. The predictive distribution map was constructed using the realization of the probabilities predicted for presence-absence of *R. pseudacacia.* The optimum site conditions have highest probabilities of species occurrence and low probabilities point to unsuitable sites for the species (Guisan & Zimmermann 2000).

### 3.4.2. Evaluation of the potential habitat model

With evaluation of the model we mean to analyse the predictive success of the model.
To estimate the predictive success of the model, the data from fieldwork on the lowland area, was used as the *training set*. Three smaller windows were set aside as *testing sets*, these were located in the hills of Goričko, lowland, and Mura floodplain.
An evaluation was done by visual measuring the adequacy between the model prediction and known *R. pseudacacia* occurrences in the *testing sites* (this can be seen in the Figures 15, 16 and 17; Guisan & Zimmermann 2000).
Then we quantified the accuracy of the predictive distribution map using three measures of accuracy: AUC, TPR and TSS. The predictions to the training site and to the testing sites were treated separately in the tests performed.

For the evaluation of the prediction to the training site a ten-fold cross-validation was chosen as a resampling technique, in which the data was randomly divided into ten roughly equal parts, and then each subset is used in turn for accuracy assessment and the remainder was used to predict those left out (Faraway 2006). This was evaluated by the area under a receiver-operating characteristic (ROC) curve (AUC), considering 0.5 as no apparent accuracy and 1.0 as perfect accuracy (Hanley & McNeil 1982). AUC is considered the standard method to assess the accuracy of predictive distribution models and it is particularly suitable for evaluation the performance of ordinal score models such as logistic regression (Allouche et al. 2006). AUC value gives a good overall estimate; how well the model reproduced the training set however this type of measure is only applicable when we have presence-absence data.

Since the absences of the species in the testing sites were not clear, due to the detectability of the species from DOF' interpretation, and also because these areas were not field surveyed, the false absences are more likely to occur than the false presences. Therefore the measure of AUC in our case can be misleading (Lobo et al. 2008). So the AUC measurement could not be applicable for the three *testing sites* since we only have presence data, thus we plotted a True Positive Rate (TPR). The TPR was also plotted for the *training site*. TPR is a metric test which concentrates in the true presences, i.e. the number of occurrences where the presence of the species was correctly identified. The TPR curve is better the higher it runs. This test was calculated over a series of cuts applied to the predicted probabilities.

Another technique used to assess the accuracy of the predictions for the *training* and *testing sites* was the True Skill Statistics (TSS). The maximum of True Skill Statistics (Allouche et al. 2006) was used to cut the probability distribution into a presence / absence binary map. TSS is a simple and intuitive measure for the performance of species distribution models when the predictions are expressed as presence-absence maps (Allouche et al. 2006). This measure it is not affected by prevalence (proportions of presences in the sample) or by the size of the specific data set used for model validation. TSS takes into account both omission and commission errors and ranges from -1 to +1, where +1 indicates perfect agreement and values equal or lower than zero indicate a random performance (Allouche et al. 2006).

Finally, the comparison of the True positive curves from the training and testing sites was done by a methods conceptually corresponding to paired-sample ANOVA. The appropriate way to perform such a test is to use linear mixed models, with site identity as random variable and the predictor variables as fixed effects (the function used in R was *lme*). A Tukey post test was also applied to the mixed model to assess pairwise significance.

# 4. Results

## 4.1. Modelling approach and prediction

The spatial relationship between the species occurrence and environmental factors was derived by using GLM.

The potential habitat model was run with six predictor variables (elevation, distance to the nearest road, distance to the nearest water body, soil type, soil quality and land use). Table 2 shows the result from the Chi-square test, which selects the variables according to its significance to the model. Elevation was dropped from the model by the variable selection, thus was not included into the model.

**Table 2:** Result from the variable selection procedure. Where Df represents the degrees of fredom; Deviance is the likelihood ratio chi-square value; and the significance p-value in the is based on the change in Deviance and the change in Df produced by adding that term into the model. Asterisks refer to significance level: '***' - p< 0.001, '**' - p<0.01, '*' - p< 0.05, '.' - p<0.1.

| | Df | Deviance | Pr(Chi) | Sign. |
|---|---|---|---|---|
| <none> | | 1450.0 | | |
| Land.Use | 5 | 1554.0 | < 2.2e-16 | *** |
| Dist_water | 1 | 1450.2 | 0.642518 | |
| Dist_road | 1 | 1455.9 | 0.015599 | * |
| Soil.type | 2 | 1490.2 | 1.863e-09 | *** |
| Soil.quality | 17 | 1570.5 | < 2.2e-16 | *** |
| Elevation | 1 | 1450.3 | 0.580862 | |
| I(Elevation^2) | 1 | 1450.3 | 0.604519 | |
| I(Dist_road^2) | 1 | 1452.9 | 0.087407 | . |
| I(Dist_water^2) | 1 | 1457.1 | 0.007834 | ** |

Table 3 shows the regression coefficients of the GLMs. Linear and quadratic terms of each continuous variable were included into the model. Significance levels show the influence of variables (if continuous) or levels (in the case of categorical variables) on model results.

**Table 3:** Regression coefficients of the GLM. The linear and quadratic terms of each predictor variable were selected when they produced a significance at a 0.05 level. Asterisks refer to significance level: '***' - p< 0.001, '**' - p<0.01, '*' - p< 0.05, '.' - p<0.1.

| Variable | Estimate value of the coefficient | p- value | Significance level |
|---|---|---|---|
| Land Use: Forest | 1.873e+01 | 0.99681 | |
| Land Use: Meadow | 2.295e+00 | 1.01e-08 | *** |
| Land Use: Orchard | 9.872e-02 | 0.94301 | |
| Land Use: Pasture | 4.934e+00 | 1.70e-07 | *** |
| Land Use: Urban | 2.133e+00 | 0.05461 | . |
| Dist. water | 6.151e-05 | 0.94773 | |
| Dist. road | -5.100e-03 | 0.02569 | * |
| Soil type: Fluvisol | -1.548e+00 | 0.00148 | ** |
| Soil type: Urban Soil | -1.782e+01 | 0.98327 | |
| Soil quality.L | 9.763e+00 | 0.99080 | |
| Soil quality.Q | -1.449e+01 | 0.98117 | |
| Soil quality.C | -1.480e+00 | 0.99779 | |
| Soil.quality^4 | 8.296e+00 | 0.99054 | |
| Soil.quality^5 | 1.055e+01 | 0.99212 | |
| Soil.quality^6 | 1.055e+01 | 0.98660 | |
| Soil.quality^7 | -1.049e+00 | 0.99831 | |
| Soil.quality^8 | 4.431e-01 | 0.99928 | |
| Soil.quality^9 | -9.550e+00 | 0.98755 | |
| Soil.quality^10 | 4.946e-01 | 0.99810 | |
| Soil.quality^11 | 1.325e+01 | 0.98479 | |
| Soil.quality^12 | -2.438e+00 | 0.99690 | |
| Soil.quality^13 | -5.986e-01 | 0.99893 | |
| Soil.quality^14 | 6.870e-01 | 0.99930 | |
| Soil.quality^15 | -5.196e+00 | 0.99374 | |
| Soil.quality^16 | -1.306e+01 | 0.98708 | |
| Soil.quality^17 | 4.125e+00 | 0.99514 | |
| I(Dist road$^2$) | 1.295e-05 | 0.09009 | . |
| I(Dist water$^2$) | -2.323e-06 | 0.04251 | * |

The result of the model shows that meadows and pastures are significantly more likely to be invaded by *R. pseudacacia* than other land use types (fields, forests, orchards and urban).

Distance from water bodies and from roads influence the occurrence of *R. pseudacacia*. Regarding the predictor soil type, the model shows that fluvisols are significantly less susceptible to *R. pseudacacia* than Cambisols and they also differ in this manner from Urban Soils.

Soil quality was also retained in the model however there is no significance of this variable to the occurrence of the species. For ordered factors the parameterization uses polynomial contrasts, the .L, .Q, and .C suffixes stand for linear, quadratic, and cubic coefficients and the exponent values of soil quality mean the number of order terms.

The influence of continuous variables (distance from water bodies and roads) in the occurrence of the study species within the study area can be seen in the Figures 12 and 13.

The distance from water bodies affects negatively the occurrence of *R. pseudacacia*, this result can be seen in Figure 12.

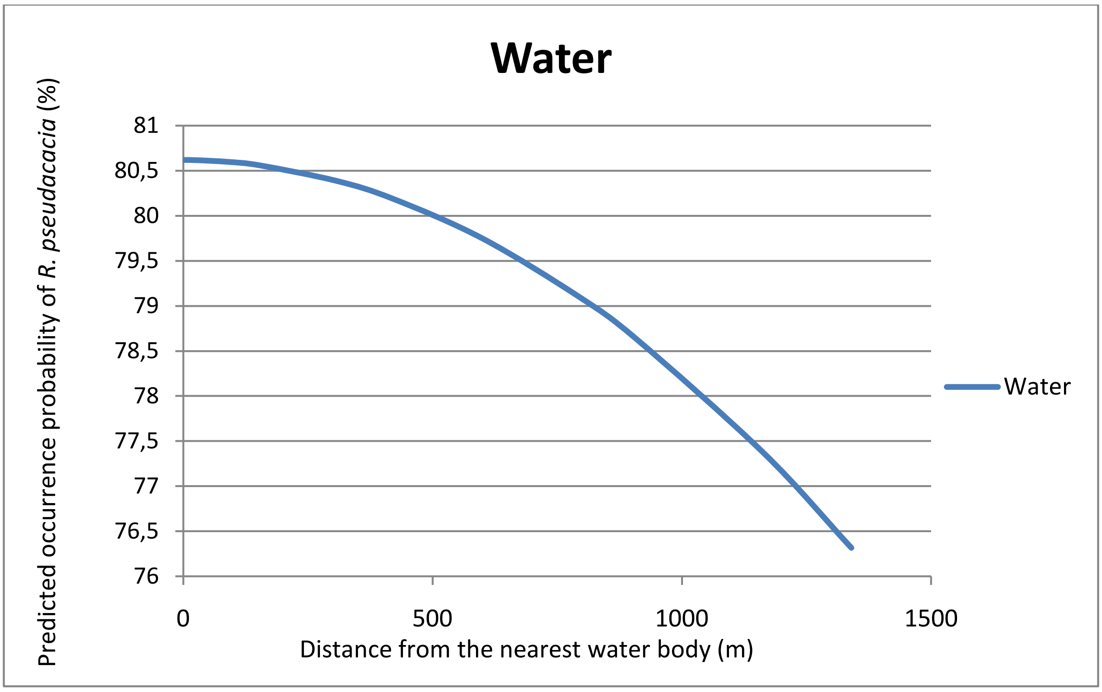


**Figure 12-** Relationship curve between distance from water and the target species fitted by GLM.

Close to the roads there is a higher probability of species occurrence, this probability decreases when the distance from road network ranges between 100-300 meters. When the distance from the road network is higher than 300m the probability of the species' occurrence increases again. The influence of this variable in the occurrence of the species in the study region can be seen Figure 13.

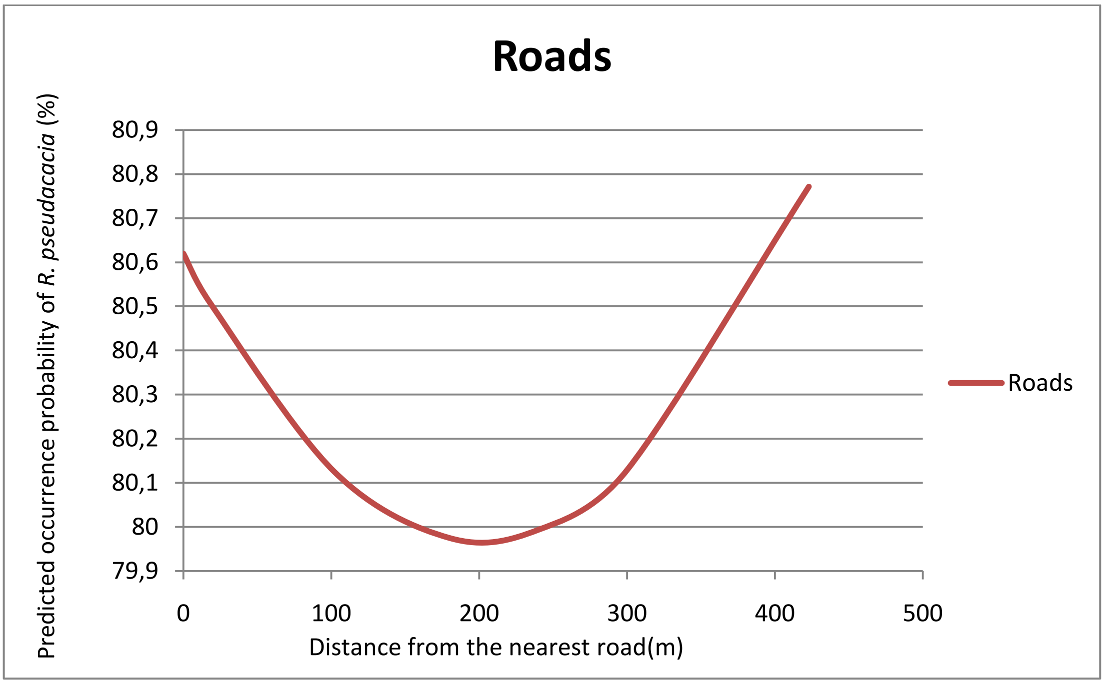


**Figure 13-** Relationship curve between distance from road and the target species fitted by GLM.

The relationship ratio between the species occurrence and the environmental factors represents the probability of *R. pseudacacia* occurrence.

Our predictions of the potential habitat map for *R. pseudacacia* were first verified by the spatial overlapping at the existing locations (for the training were obtained by field survey and for the testing sites were obtained by DOF interpretation) with the predicted values of probability occurrence. We assumed that probabilities of less than 0.30 represent areas where the species is unlikely to occur (Zimmermann & Kienast 1999).

The model generated an overall satisfactory agreement between “simulated” *R. pseudacacia* and “observed” in the *training* and *testing sites*.

Figure 14 represents the potential habitat map for *R. pseudacacia* for the *training site*, in which a probability of occurrence was assigned to each sampling point. The indices of the potential habitat map were sorted in ascending order into five classes with approximately values of: 0.00-0.06 (dark green colour); 0.06-0.15 (light green colour); 0.15-0.37 (yellow colour); 0.37-0.68 (orange colour) and 0.68-1.00 (red colour).

A very good agreement between observed and simulated *R. pseudacacia* distribution was achieved for the *training site*. The predictions resulting from the predictive model correlate well to the field observations; most observations fall into the predicted category with a higher probability than 0.68. Some field observations which were outside this probability were included in the next category (0.37-0.68) (see Figure 14).


**Figure 14-** Probability of occurrence for R. pseudacacia in the training site.

Figures 15, 16 and 17, represent the habitat potential map for *R. pseudacacia* for the testing sites located on lowland, floodplain of Mura River and Goričko, respectively, in which a probability of occurrence was assigned to each sampling point.

Predicted habitat distribution for the *testing sites* where the species is not present, infer that those sites are suitable to be invaded. The indices of the potential habitat map were sorted into five classes, however these classes have different ranks for each testing sites. For the lowland testing site the indices of the potential habitat map were classed into the follow approximately values: 0.00-0.06 (dark green colour); 0.06-0.13 (light green colour); 0.13-0.29 (yellow colour); 0.29-0.54 (orange colour) and 0.54-1.00 (red colour).

A fair agreement between observed and simulated *R. pseudacacia* distribution was achieved for lowland *testing site* as can be seen in Figure 15.


**Figure 15-** Occurrence probabilities of *R. pseudacacia* for the lowland testing site as predicted by GLM using environmental predictors.

For the floodplains of Mura River testing site the indices of the potential habitat map were classed into the follow approximately values: 0.00-0.08 (dark green colour); 0.08-0.20 (light green colour); 0.20-0.35 (yellow colour); 0.35-0.78 (orange colour) and 0.78-1.00 (red colour).

A good agreement between observed and simulated *R. pseudacacia* distribution was achieved for the Mura floodplain *testing site* as can be seen in Figure 16.

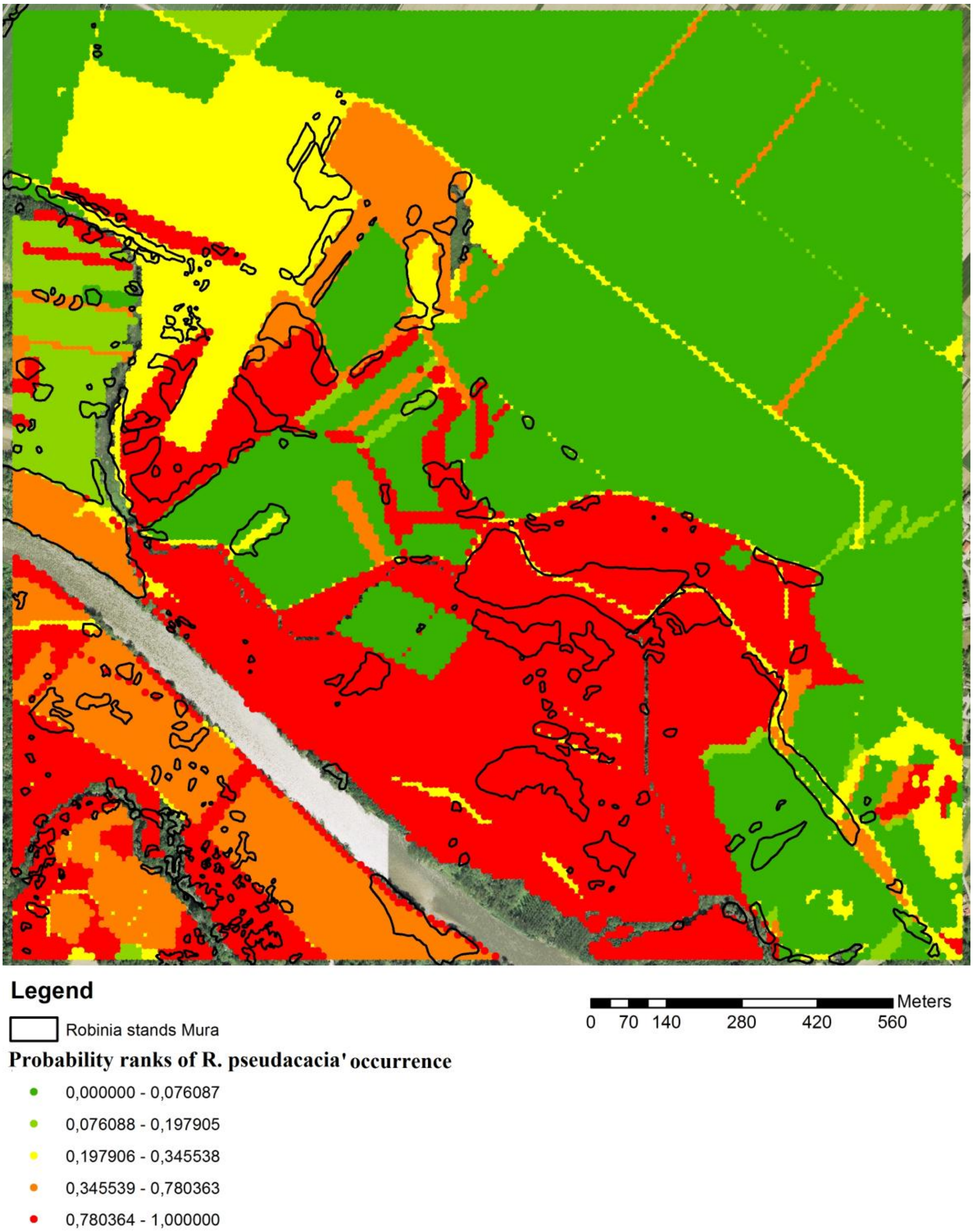


**Figure 16-** Occurrence probabilities of *R. pseudacacia* for the Mura floodplains testing site as predicted by GLM using environmental predictors.

For the Goričko testing site the indices of the potential habitat map were classed into the follow approximately values: 0.06-0.15 (dark green colour); 0.15-0.29 (light green colour); 0.29-0.62 (yellow colour); 0.62-0.86 (orange colour) and 0.86-1.00 (red colour).

A good agreement between observed and simulated *R. pseudacacia* distribution was achieved for Goričko *testing site* as can be seen in Figure 17.

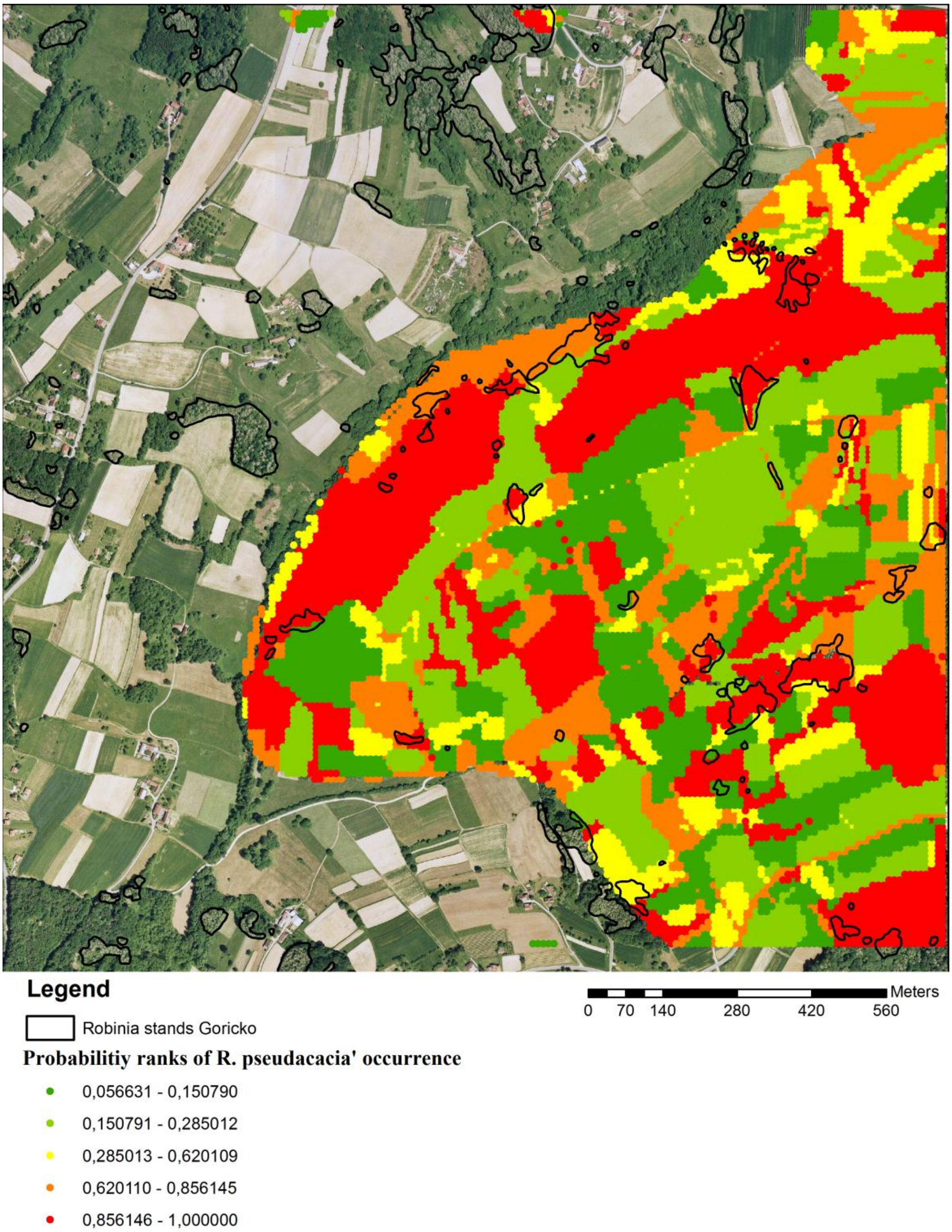


**Figure 17-** Occurrence probabilities of *R. pseudacacia* for the Goričko testing site as predicted by GLM using environmental predictors.

## 4.2. Evaluation of the potential habitat model

As previously mentioned, we evaluated the predictive distribution map using three measures of accuracy: AUC, TPR and TSS. For the testing sites we did not calculated AUC as an evaluation measure of predictive model since AUC is a synthetic statistic and is used if we are equally sure in presences and absences in prediction, what is not the case for these sites. The other two measures were applied to both training and testing sites.

The median performance of the best model for the *training site* was an AUC value of 0.8911.

Figure 18 shows the result of TPR for the training site, by observation of this plot we can see that around 80% of the observed occurrences were predicted with more than 60% of probability of occurrence.

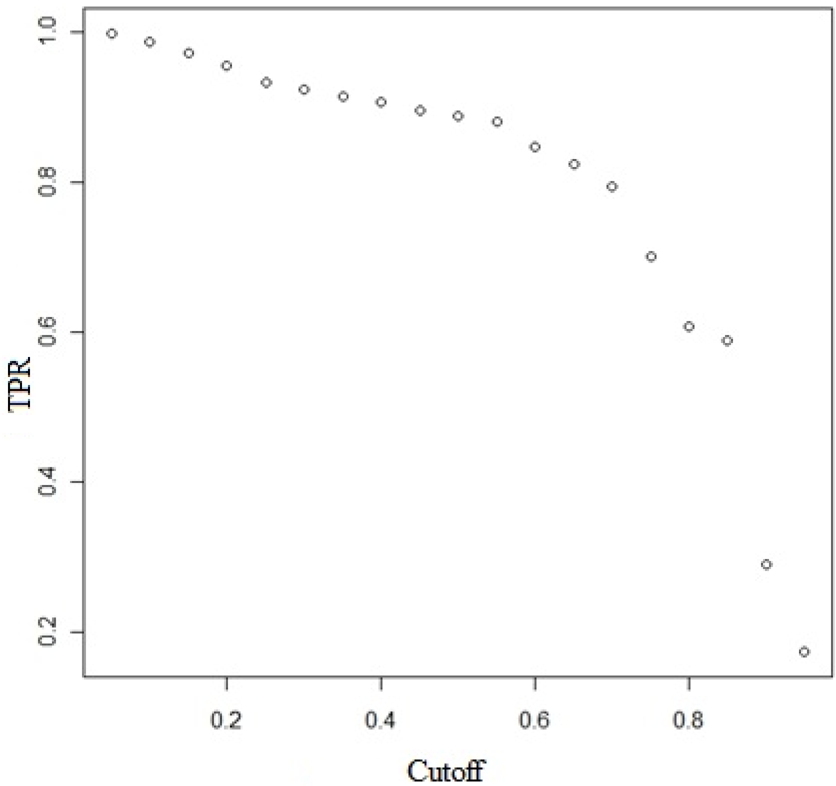


**Figure 18-** Plot of the True Positive Rate result from the *training site.*

The results of TPR for the three testing sites were plotted together (Figure 19) in order to make comparisons between the fit of the model in these three sites.

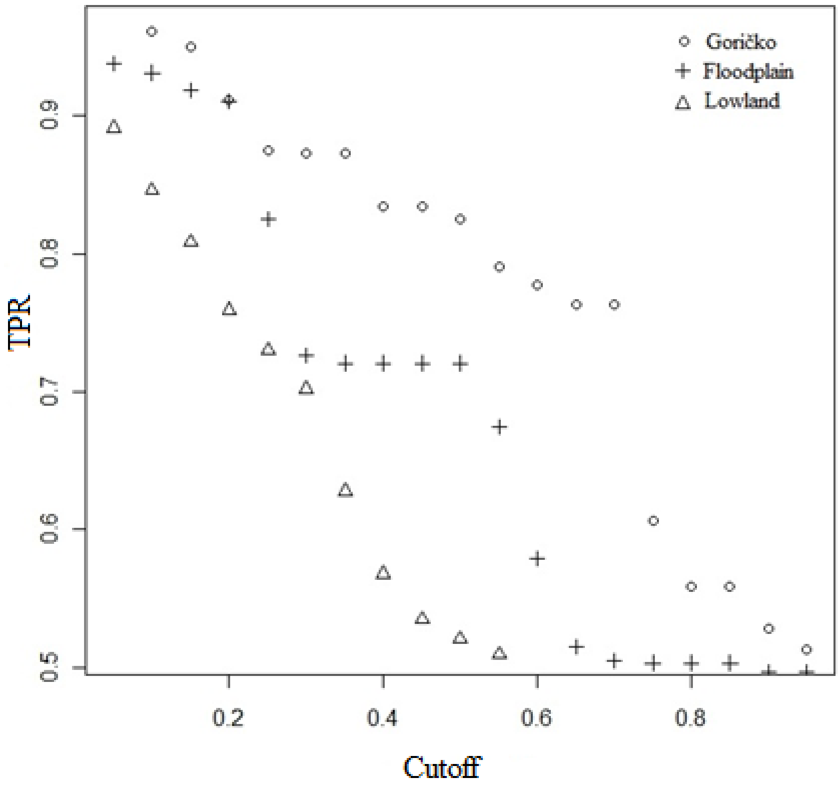


**Figure 19-** Plot of the True Positive Rate results for the three *testing sites.*

As it can be seen in the Figure 19, TPR shows that the model performed the best for the Goričko *testing site* followed by Mura floodplain *testing site* and did not perform very well for the lowland *testing site*.

The result of cut-offs from TSS for the *training site* can be seen in the Table 4, and for the lowland, Mura floodplain and Goričko testing sites these results can be seen in the Tables 5, 6, and 7, respectively.

**Table 4:** Accuracy measurement result for the training site.

| | **cut** | **TPR** | **TSS** |
|---|---|---|---|
| 1 | 0.05 | 0.9986941 | 0.2218738 |
| 2 | 0.10 | 0.9869953 | 0.3871460 |
| 3 | 0.15 | 0.9730928 | 0.5119185 |
| 4 | 0.20 | 0.9560072 | 0.6173997 |
| 5 | 0.25 | 0.9333714 | 0.6564659 |
| 6 | 0.30 | 0.9239852 | 0.6595535 |
| 7 | 0.35 | 0.9145718 | 0.6568316 |
| 8 | 0.40 | 0.9069268 | 0.6690563 |
| 9 | 0.45 | 0.8961258 | 0.6770800 |
| 10 | 0.50 | 0.8877462 | 0.6868698 |
| 11 | 0.55 | 0.8803189 | 0.6882327 |
| 12 | 0.60 | 0.8477527 | 0.6681403 |
| 13 | 0.65 | 0.8237295 | 0.6648787 |
| 14 | 0.70 | 0.7952987 | 0.6622269 |
| 15 | 0.75 | 0.7018718 | 0.6064956 |
| 16 | 0.80 | 0.6076015 | 0.5489768 |
| 17 | 0.85 | 0.5893188 | 0.5387020 |
| 18 | 0.90 | 0.2892045 | 0.2659601 |
| 19 | 0.95 | 0.1736043 | 0.1645052 |

**Table 5:** Accuracy measurement result for the lowland testing site.

| | **cut** | **TPR** | **TSS** |
|---|---|---|---|
| 1 | 0.05 | 0.89123196 | 0.35255011 |
| 2 | 0.10 | 0.84572697 | 0.55485445 |
| 3 | 0.15 | 0.80854606 | 0.59418485 |
| 4 | 0.20 | 0.75915649 | 0.61591942 |
| 5 | 0.25 | 0.72974473 | 0.60113482 |
| 6 | 0.30 | 0.70199778 | 0.57795673 |
| 7 | 0.35 | 0.62763596 | 0.50857292 |
| 8 | 0.40 | 0.56770255 | 0.45467449 |
| 9 | 0.45 | 0.53496115 | 0.42786579 |
| 10 | 0.50 | 0.52053274 | 0.41851767 |
| 11 | 0.55 | 0.50943396 | 0.41297653 |
| 12 | 0.60 | 0.46226415 | 0.39444731 |
| 13 | 0.65 | 0.31687014 | 0.28134217 |
| 14 | 0.70 | 0.22031077 | 0.20108065 |
| 15 | 0.75 | 0.16037736 | 0.14742090 |
| 16 | 0.80 | 0.13817980 | 0.12815559 |
| 17 | 0.85 | 0.10710322 | 0.09759045 |
| 18 | 0.90 | 0.09322974 | 0.08378517 |
| 19 | 0.95 | 0.09322974 | 0.08381926 |

**Table 6:** Accuracy measurement result for the Mura floodplains testing site.

| | **cut** | **TPR** | **TSS** |
|---|---|---|---|
| 1 | 0.05 | 0.9382812 | 0.4246196 |
| 2 | 0.10 | 0.9312500 | 0.4340655 |
| 3 | 0.15 | 0.9183594 | 0.4497074 |
| 4 | 0.20 | 0.9101562 | 0.4506120 |
| 5 | 0.25 | 0.8250000 | 0.3967471 |
| 6 | 0.30 | 0.7265625 | 0.3750263 |
| 7 | 0.35 | 0.7207031 | 0.3756292 |
| 8 | 0.40 | 0.7207031 | 0.3756292 |
| 9 | 0.45 | 0.7207031 | 0.3757426 |
| 10 | 0.50 | 0.7207031 | 0.3765362 |
| 11 | 0.55 | 0.6746094 | 0.3381897 |
| 12 | 0.60 | 0.5792969 | 0.2791193 |
| 13 | 0.65 | 0.5144531 | 0.2351742 |
| 14 | 0.70 | 0.5042969 | 0.2414950 |
| 15 | 0.75 | 0.5031250 | 0.2501489 |
| 16 | 0.80 | 0.5031250 | 0.2523030 |
| 17 | 0.85 | 0.5031250 | 0.2532856 |
| 18 | 0.90 | 0.4960938 | 0.2509783 |
| 19 | 0.95 | 0.4960938 | 0.2509783 |

**Table 7:** Accuracy measurement result for the Goričko testing site.

| | **cut** | **TPR** | **TSS** |
|---|---|---|---|
| 1 | 0.05 | NA | NA |
| 2 | 0.10 | 0.9613971 | 0.04581939 |
| 3 | 0.15 | 0.9503676 | 0.14200709 |
| 4 | 0.20 | 0.9117647 | 0.24381104 |
| 5 | 0.25 | 0.8750000 | 0.30193793 |
| 6 | 0.30 | 0.8731618 | 0.31123720 |
| 7 | 0.35 | 0.8731618 | 0.31123720 |
| 8 | 0.40 | 0.8345588 | 0.29312728 |
| 9 | 0.45 | 0.8345588 | 0.30292829 |
| 10 | 0.50 | 0.8253676 | 0.30168187 |
| 11 | 0.55 | 0.7904412 | 0.28873342 |
| 12 | 0.60 | 0.7775735 | 0.30757056 |
| 13 | 0.65 | 0.7628676 | 0.31996595 |
| 14 | 0.70 | 0.7628676 | 0.34469123 |
| 15 | 0.75 | 0.6066176 | 0.25578605 |
| 16 | 0.80 | 0.5588235 | 0.25766523 |
| 17 | 0.85 | 0.5588235 | 0.26152623 |
| 18 | 0.90 | 0.5275735 | 0.26005051 |
| 19 | 0.95 | 0.5128676 | 0.26776815 |

The pairwise comparison results indicate that the majority of presences predicted fell into the presences observed for the training and testing sites, these results can be seen from the tables 8 to 11.

**Table 8:** Pairwise comparison matrix for the presence /absence [1/0] of *R. pseudacacia* observed and predicted for the *training site* according to TSS.

| Predicted Robinia | Observed Robinia | |
|---|---|---|
| | **0** | **1** |
| **0** | 279867 | 4399 |
| **1** | 66540 | 32357 |

**Table 9:** Pairwise comparison matrix for the presence /absence [1/0] of *R. pseudacacia* observed and predicted for the lowland *testing site* according to TSS.

| Predicted Robinia | Observed Robinia | |
|---|---|---|
| | **0** | **1** |
| **0** | 25128 | 434 |
| **1** | 4201 | 1368 |

**Table 10:** Pairwise comparison matrix for the presence /absence [1/0] of *R. pseudacacia* observed and predicted for the Mura floodplain *testing site* according to TSS.

| Predicted Robinia | Observed Robinia | |
|---|---|---|
| | **0** | **1** |
| **0** | 14301 | 230 |
| **1** | 12160 | 2330 |

**Table 11:** Pairwise comparison matrix for the presence /absence [1/0] of *R. pseudacacia* observed and predicted for the Goričko *testing site* according to TSS.

| Predicted Robinia | Observed Robinia | |
|---|---|---|
| | **0** | **1** |
| **0** | 7836 | 129 |
| **1** | 5632 | 415 |

The subsequent Tukey post test demonstrated, as seen in Table 12, a marginal significance between the predictions for the *training site* and Goričko *testing site*, being the last a bit worse predicted than the former. While the comparison between the predictions for the *training site* and *Mura floodplain* and *Lowland testing sites* shows a significantly difference, so these *testing sites* are significantly worse predicted than the *training site*.

**Table 12:** Comparison of TPR curves by paired-sample ANOVA with Tukey post test

Asterisks refer to significance level: '***' - p< 0.001, '**' - p<0.01, '*' - p< 0.05, '.' - p<0.1.

| | **Estimate** | **Std. Error** | **z value** | **Pr(>\|z\|)** | **Sign.** |
|---|---|---|---|---|---|
| Lowland - Goričko == 0 | 0.018627 | 0.003234 | 5.760 | < 0.001 | *** |
| Mura - Goričko == 0 | 0.026104 | 0.003234 | 8.073 | < 0.001 | *** |
| Training - Goričko == 0 | 0.007799 | 0.003234 | 2.412 | 0.0748 | . |
| Mura - Lowland == 0 | 0.007477 | 0.003177 | 2.353 | 0.0861 | . |
| Training - Lowland == 0 | -0.010828 | 0.003177 | -3.408 | 0.0038 | ** |
| Training - Mura == 0 | -0.018305 | 0.003177 | -5.761 | < 0.001 | *** |

# 5. Discussion

The present study examined the relationship between environmental factors and the distribution of *Robinia pseudacacia* in Prekmurje region, this relationship was established by GLM from the binomial family.

## 5.1. Are land use and other environmental variables co-acting in determining the current spatial distribution of *R. pseudacacia* in Prekmurje region?

To answer to this question we collected information of 6 environmental variables (elevation, land use, soil quality and soil type, distances from the nearest water body and road), which we believed that directly or indirectly influence the occurrence of this species in the study region. To establish the relationship between *R. pseudacacia* occurrences and environmental factors we performed a logistic regression model (GLM with binomial family). The result of the model showed that 5 out of 6 environmental factors may explain the current occurrence of the species in the study region. Elevation was eliminated from the model based on the chi-square selection criterion. This variable, contrarily to what has been shown by other studies (Higgins et al. 1999; Pauchard & Alaback 2004), does not play a role in the distribution of studied species in Prekmurje. Nevertheless we have to note that in our study area the differences in topographical features such as elevation are small (see Table 1) so this result could be expected.

The remaining environmental factors were kept in the calibrated model (see Table 3), since they appear to have some influence on the distribution of the species.

The most influential predictor in the model was land use; this factor had the highest predictive power. Several modelling studies (Pyšek et al. 1998; Pauchard & Alaback 2004; Mosher et al. 2009) also found that land use is an important explanatory variable for predicting invasive species distribution.

From the land use types, the effect of fields was not significant in the model, we suggest that areas with highest agricultural potential as fields are still maintained for agricultural purposes so they do not contribute to *R. pseudacacia* occurrence, neither naturally nor planted.

Pastures and meadows are the land use types more prone to be invaded by *R. pseudacacia*, this result corroborates also with other studies (Boring & Swank 1984) which defends that these areas are often rich in plant invaders largely due to anthropogenic action (Heywood 1989). Usually pastures and meadows are the first surfaces to be abandoned by farmers, when we compare with agricultural fields, if this is also true for Prekmurje region, this could explain the occurrence of the species in these surfaces.

The effect of urban land use was only marginally significant for the potential distribution of the species in the study area.

By the visualization of the Figure 12 we can see that *R. pseudacacia* appears closest to water bodies. The distance from water bodies has a negative influence on the species occurrence in Prekmurje. This result corroborates with what was showed by Rudolf & Brus (2006), where the species occurs in areas with high ground water level and stagnant water, this is possible because those surfaces are settled in well drained soils. Also, in its native range the species appears to have affinity to moister soils (Huntley 1990). Although, several studies (Bartha et al. 2008; Kleinbauer et al. 2010; Pyšek et al. 1998; Walter et al. 2005) have demonstrated that *R. pseudacacia* appears in dry or semi-dry habitats.

We can say that the road network influence the distribution of the species, this result can be seen in the Figure 13. This predictor influences *R. pseudacacia* distribution in an interesting way; closer to the roads the probability of finding the species is higher, this probability decreases when the roads are within a distance between 100-300 m, and a higher distance than 300m increases again the probability of *R. pseudacacia* occurrence. It is true that the species is frequently planted along roads (Bartha et al. 2008); this can explained the occurrence of the species in distances shorter than 100m.

We believe that the decreasing of probability of the species' occurrences within a distance between 100- 300m may be due to the existence of agricultural surfaces in those areas due to the easiest accessibility. However the distances between agricultural surfaces and roads were not measured in this study so we can just infer. With the distance of at least 300m from the nearest road, the probability *R. pseudacacia* occurrence increases, if we assume that the further areas are the first areas to be abandoned by farmers this result could have a support.

## 5.2. Can we predict a potential habitat of *R. pseudacacia* for other areas based on the current habitat of the species in a small fragment of the investigation area?

The model used to explain the patterns of distribution of *R. pseudacacia* was based on the current locations of the species on the training site. This model was evaluated by the AUC value, and the median performance of the best model for the *training site* was an AUC value of 0.8911, this indicated spatial agreement between the model prediction and actual *R. pseudacacia* sites from the training data. Many authors agree that a performance between 0.7 and 0.9 of AUC values corresponds to high model performance (Guisan et al. 2007). The cross-validation performed shows that we have created a robust model. Since the model was robust we applied it to predict the potential habitat of the species in three testing sites. This prediction was evaluated first by the overlapping of the simulated and observed (from DOF' interpretation) *R. pseudacacia*. Then we performed the TPR and TSS tests to evaluate the accuracy of the predicted potential distribution map of the species. These tests showed that the model performed better for Goričko *testing site*, followed by Mura floodplain *testing site* and finally for the lowland *testing site* (this result can be seen by the Figure 19). However we did expect that the model would have a better fit in the lowland *testing site* since this area is ecologically and geographically more similar to the *training site* than the other two *testing sites*. This result may be due to several reasons. One of them is that the digital soil map used may be not fine enough for small scale studies. We checked some areas to understand the result and the areas where the model predicted worse the occurrence of the species were considered as Urban soil in the digital soil map, although these areas were assigned as different land use than urban in the land use data. We made a comparison between two parcels and both parcels had the similar environmental characteristics, regarding the environmental variables used in the analysis, the only difference between these two parcels was that one was considered as urban soil and the other as cambisol. This difference was enough to obtain different probability of *R. pseudacacia* occurrence as can be seen by the Table 13.

**Table 13 :** Comparison between *R. pseudacacia* stand divided by the limit of soil type classification.

| **Robinia presence** | 1 | 1 |
|---|---|---|
| **Prediction of occurrence** | 0.76 | 0 |
| **Distance to roads** | 38m | 46m |
| **Distance to water** | 241m | 236m |
| **Land use type** | Meadow | Meadow |
| **Soil type** | Cambisol | Urban soil |
| **Soil quality** | 34 | 34 |

As could be seen in the Figure 17 Goričko *testing site* was not fully predicted, this because the non-predicted area belongs to a different soil type that was not used in the model building. Since the predictive model is only useful for similar environmental characteristics as the training data we eliminated the area of Goričko that belonged to another soil type. This is an inherent limitation of the approach used as is showing up here.

As the occurrence, and therefore, spatial distribution of *R. pseudacacia* in the study area is also human induced (Rudolf & Brus 2006), we believe that the addition of social explanatory variables, as predictors of species occurrence, could improve the model fit, thus improve the accuracy of the potential distribution of the species in the testing sites. Since the distribution of an invasive species, as *R. pseudacacia*, is determined by the combination of environmental factors and the dispersal process of the species (D'Antonio et al. 2001), and we did not incorporate the propagule dispersal of *R. pseudacacia* into our potential habitat model we may have here a weakness in our model.

Notwithstanding we believe that the model based on the current distribution of the *R. pseudacacia* may serve as a valuable tool to predict the potential habitat of the species in the testing sites.

## 5.3. Modelling approach

The application of a GLM to predict the potential habitat (or ecological niche) of invasive species is not new (Hortal et al. 2010) and has been tested and compared with other distribution models (Guisan & Zimmerman 2000; Guisan et al. 2007; Mingyang et al. 2008). We choose GLM with a logistic link function as a tool for the modelling procedure due to several reasons: GLM constitute a more flexible family of regression models which allow the binomial distribution, and our dependent variable has a binomial response (0/1); it yielded the prediction between the limits of observed values (0/1); allowed to include higher polynomials of the continuous variables; predicted the potential distribution of *R. pseudacacia* from the species response; and it was easy to implement the GLM result into a GIS environment, which resulted in the potential habitat maps of *R. pseudacacia*.

We believe that GLM is an appropriate method for modelling the potential distribution of *R. pseudacacia*.

# 6. Conclusions

We studied the distribution patterns of the invasive species *R. pseudacacia* in the Prekmurje region. For this we analyzed the influence of environmental factors in the species occurrences by using GLM. With the calibrated model we predicted the potential habitat of the species for three testing sites located on the hilly area of Goričko, lowland area and Mura River floodplain.

The results from the model indicated the most important environmental factors for the species occurrence; and these are land use, soil type, soil quality, distance to the nearest road, and distance to the nearest water body. The major part of the observed distribution pattern was explained by land use, being meadows and pastures the most prone to be invaded by *R. pseudacacia*.

We conclude that environmental factors together with land use information can explain the occurrence and distribution of the species in the study area, however *R. pseudacacia* was intentionally introduced into the region and it's still being planted by farmers and beekeepers what makes the process of invasion also human-induced.

The logistic regression (GLM with logit link) was appropriate to find the most influential environmental variables and model the potential habitat of the species. From the current species distribution we could predict the potential habitats for three sites.

Building models of real world processes is an attempt to predict how these processes operate under certain conditions, but cannot explain all the mechanisms.

Raising awareness among the public about this invasive species is vital to combat this environmental problem, especially in the study region, where only the benefits of the species are taken into account.

## 6.1. Future work

Future work should develop a simultaneous model that integrates both attributes of the invasion process, which are the environmental factors and dispersal process, to estimate a better potential habitat for *R. pseudacacia*.

It is important to note that the species was introduced and it's still planted in the region, thus we believe that the model presented in this work could be improved by adding social factors as predictor variables.

Though the model performed adequately, the predicted potential distribution of *R. pseudacacia* in Prekmurje region, its generalizability to other regions may be limited. In order to increase its generalizability we should analyze the distribution of the species in these regions and maybe include other environmental factors, which did not play an important role in our study region but it may be important for other regions.

Climate factors at the small scale, as the regional scale, may not play so important role for the species distribution but at a bigger scale these factors can be consider as dominant factors.

## 7. Summary

Our study presents the distribution patterns of the invasive species *R. pseudacacia* in Prekmurje, in the Northeastern Slovenia. We established two hypotheses: H1) Land use and other environmental variables co-act in determining the current spatial distribution of *R. pseudacacia* in Prekmurje region; and H2) Based on the current habitat of the species in a small fragment of the investigation area we can predict distribution of *R. pseudacacia* for other areas.

We selected environmental variables that were likely to influence directly and indirectly the occurrence of the species in the study region. These variables were: elevation, land use, soil type, soil quality, distance to the nearest road and distance to the nearest water body. The environmental variables were used as predictors and the dependent variable was the *R. pseudacacia* occurrence, the last was a binomial variable which ranges between 0-1 (where 0 represents the absence and 1 represents the presence of the species). A Generalized linear model of the binomial family was used to investigate the effects of environmental variables and to predict the spatial distribution of *R. pseudacacia* in three *testing sites* settled outside the range of the *training site* (data used to build the model). The environmental variables that showed significance for the model, thus critical influencing factors for the current occurrence of the species within the study area are: land use, soil type, distance to the nearest road, and distance to the nearest water body. Elevation was dropped out from the model; this variable does not influence the occurrence of the species in Prekmurje region.

With the calibrated model the potential distribution of the study species for three *testing sites* (Goričko, lowland, and Mura floodplain) was predicted.

To evaluate the model fit three measures we performed: AUC, TPR and TSS.

The model fitted better in the Goričko *testing site*, followed by the Mura floodplain *testing site* and finally the lowland *testing site*, contrarily to what was expected. We expected that the model would have the best fit in the lowland *testing site* since it has similar ecological and geographical features as the *training site*. However, we think that the results of our model were satisfactory.

Since the occurrence and spatial distribution of *R. pseudacacia* in the study area is human induced, we believe that the inclusion of additional explanatory variables, e.g. social factors, could improve the model fit. Future work should develop a simultaneous

model that integrates the environmental factors and dispersal process, to estimate a better potential habitat for *R. pseudacacia*.

## 8. References

Allan, H.H. (1961). Flora of New Zealand. Volume I, Indigenous Tracheophyta. Wellington: R.E. Owen Government Printer.

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

ARSO- Agencija Republike Slovenije za Okolje. URL: http:// arso.gov.si/. Consulted on April 4, 2011.

Ball, P. W. (1968). 30.Robinia L. In: Edited by T. G. Tutin, V. H. Heywood, N. A. Burges, D. M. Moore, D. H. Valentine, S. M. Walters and D. A. Webb. Flora Europaea, Vol. 2, pp.106. Cambridge University Press, Great Britain.

Bartha, D., Csiszár, Á. & Zsigmond, V. (2008). Black Locust (Robinia pseudoacacia L.). In: Edited by Zoltán Botta-Dukát and Lajos Balogh. The most important invasive species in Hungary. Institute of Ecology and Botany, Hungarian Academy of Sciences, Vácrátót, Hungary, pp. 63-76.

Boring, L. R. & Swank, W. T. (1984). The role of black locust (Robinia pseudoacacia) in forest succession. Journal of Ecology 72: 749-766.

Christen, D. & Matlack, G. (2006). The role of roadsides in plant invasions: a demographic approach. Conservation Biology 20: 385-391.

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton: Princeton University Press, p282.

Cunder, T. (2009). Kmetijstvo v Pomurju danes in jutri. In: Pomurje: Trajnostni regionalni razvoj ob reki Muri. Zbornik povzetkov. Ljutomer, Murska Sobota.

Čarni, A., Košir, P., Marinček, L., Marinšek, A., Šilc, U. & Zelnik, I. (2008). Commentary to the Vegetation Map of Forest Communities of Slovenia in Scale of 1:50000- Section Murska Sobota. PAZU, Murska Sobota.

DAISIE. (2006). Delivering Alien Invasive Species Inventories for Europe: Robinia pseudoacacia. URL: http://www.europe-aliens.org/. Consulted on April 18, 2011.

D'Antonio, C. M. & Vitousek, P. M. (1992). Biological invasions by exotic grasses, the grass-fire cycle, and global change. Annual Review of Ecology and Systematics 23: 63-87.

D'Antonio, C. M. & M. C. Mack. (2001). Exotic grasses slow invasion of a nitrogen fixing tree into a Hawaiian woodland. Biological Invasions 3: 69-73.

Duchaufour, P. (1998). Handbook of Pedology: Soils, Vegetation, Environment. A. A. Balkema, Rotterdam, Brookfield.

Dukes, J. S. & Mooney, H. A. (1999). Does global change increase the success of biological invaders? Trends in Ecology and Evolution 14: 135-139.

Dullinger, S., Kleinbauer, I., Peterseil, J., Smolik, F. & Essl, F. (2009) Niche based species distribution modelling of an invasive alien plant: effects of population status, propagule pressure, and invasion history. Biological Invasions 11: 2401–2414.

Essl, F., Dullinger, S., Rabitch, W., Hulme, P. E., Hülber, K., Jarošik, V., Kleinbauer, I., Krausmann, F., Kühn, I., Nentwig, W., Vilà, M., Genovesi, P., Gherardi, F., Desprez-Loustau, M. L., Roques, A. & Pyšek, P. (2011). Socioeconomic legacy yields an invasion debt. PNAS 108: 203-207.

FAO- Food and Agriculture Organization. URL: http://www.fao.org/. Consulted on March 10, 2011.

Faraway, J. J. (2004). Linear Models with R. Boca Raton, FL: Chapman & Hall/CRC.

Faraway, J. J. (2006). Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Boca Raton, FL: Chapman & Hall/CRC.

Field, A. (2009). Discovering Statistics Using SPSS (3rd Ed.). SAGE Publications Ldt. London.

Forman, R. T. T. (1995). Land Mosaics: The ecology of landscapes and regions. Cambridge University Press, New York.

Franklin, J. (1995). Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography 19: 474-499.

Fukasawa, K., Koike, F. , Tanaka, N. & Otsu, K. (2009). Predicting future invasion of an alien tree in a Japanese oceanic island by process-based statistical models using recent distribution maps. Ecological research 24: 965-975.

Gabrovec, M. & Kladnik, D. (1997). Some new aspects of land use in Slovenia. Geografski zbornik 37: 7-64.

Gillham, J. H., Hild, A. L., Johnson, J. H., Hunt, E. R. JR. & Whitson, T. D. (2004). Weed invasion susceptibility prediction model for use with geographic information systems. Arid Land Research and Management 18: 1-12.

Globevnik, L. & Kaligarič, M. (2005). Hydrological changes of the Mura River in Slovenia, accompanied with habitat deterioration in riverine space. RMZ- Materials and Geoenvironment 52: 45-49.

Groves, R.H. (1986). Invasion of Mediterranean ecosystems by weeds. In: B. Dell, A.J.M. Hopkins and B.B. Lamont (Eds.), Resilience in Mediterranean-type Ecosystems.

Junk, Henderson, L. 1991. Alien invasive Salix spp. (willows) in the grassland biome of South Africa. South African Forestry Journal 157: 91-95.

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

Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993-1009.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C. C., Aspinall, R. & Hastie, T. (2006). Making better biogeographical predictions of species' distributions. Journal of Applied Ecology 43: 386-392.

Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Philips, S. & Peterson, A. T. (2007). What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? Ecological Monographs 77: 615-630.

Gurevitch, J. & Padilla, D. K. (2004). Are invasive species a major cause of extinctions? Trends in Ecology and Evolution 19: 470-474.

Gurevitch, J., Scheiner, S. M. & Fox, G. A. (2006). The Ecology of Plants (2nd Ed.). Sinauer Associates, Inc., USA.

Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143: 29-36.

Hastie, T., Tibshirani, R. & Freidman, J. (2009). The Elements of Statistical Learning: Data mining, Inference, and Prediction (2nd Ed.). Springer.

Heywood, V. H. (1989). Patterns, Extents and Modes of Invasions by Terrestrial Plants. In: Biological Invasions: a Global Perspective, pp. 31-60. John Wiley & Sons ltd.

Hierro, J. L., Maron, J. L. & Callaway, R. M. (2005). A biogeographical approach to plant invasions: the importance of studying exotics in their introduced and native range. Journal of Ecology 93: 5-15.

Higgins, S. I., Richardson, D. M., Cowling, R. M. & Trinder-Smith, T. H. (1999). Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. Conservation Biology 13: 303–313.

Holcombe, T. R., Stohlgren, T. J. & Jarnevich, C. S. (2010). From points to forecasts: predicting invasive species habitat suitability in the near term. Diversity 2: 738-767.

Hortal, J., Borges, P. A. V., Jiménez-Valverde, A., Azevedo, E. B. & Silva, L. (2010). Assessing the areas under risk of invasion within islands through potential distribution modelling: The case of *Pittosporum undulatum* in São Miguel, Azores. Journal for Nature Conservation 18: 247-257.

Hruška, K. (1991). Human impact on the forest vegetation in the western part of the Pannonic Plain (Yugoslavia). Vegetation 92: 161-166.

Huenneke, L. F., Hamburg, S. P., Koide, R., Mooney, H. A. & Vitousek, P. M. (1990). Effects of soil resources on plant invasion and community structure in Californian serpentine grassland. Ecology 71: 478-491.

Huntley, J. C. (1990). Robinia pseudacacia L. (black locust). In: Burns, R. M., Honkala, B. H. (eds.), Silvics of North America, vol. 2. Hardwoods Agricultural Handbook. Washington, pp. 755-761.

IUSS Working Group WRB. (2007). World Reference Base for Soil Resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome.

Jogan, N. (2000). Neofiti- rastline pritepenke. Proteus 63: 31-36.

Jogan, N. (2001). Gradivo za Atlas flore Slovenije. Center za kartografijo favne in flore, Miklavž na Dravskem polju, 443 pp.

Jogan, N. (2009). Tujerodne rastline v Sloveniji. In: Veevliet, J. K. Tujerodne vrste: Priročnik za narovovarstvenike. Zavod za Symbiosis, Grahovo, pp. 14-16.

Johnson, J. H., Hammerlink, J. D. & Gillham, J. H. (2001). An ArcView-based application for the management of noxious weed species in Wyoming. 2001 International ESRI User Conference. July, San Diego, CA. Paper.

Kalwij, J. M., Milton, S. J. & McGeoch, M. A. (2008). Roads verges as invasion corridors? A spatial hierarchical test in an arid ecosystem. Landscape Ecology 23: 439-451.

Kleinbauer, I., Dullinger, S., Peterseil, J. & Essl, F. (2010). Climate change might drive the invasive tree *Robinia pseudacacia* into nature reserves and endangered habitats. Biological Conservation 143: 382-390.

Kleinbaum, D. G., Kupper, L. L., Nizam, A. & Muller, K. E. (2008). Applied Regression Analysis and Other Multivariable Methods (4th edition). Duxbury Press.

Kowarik, I. (1995) On the role of alien species in urban flora and vegetation. In: Pyšek, P., Prach, K., Rejmánek, M. & Wade, M. (Eds) Plant invasions-general aspects and special problems. SPB Academic Publishing, Amsterdam, pp 85–103.

Kowarik, I. (2003). Biologische Invasionen- Neophyten und Neozoen in Mitteleuropa. E. Ulmer Verlag, Stuttgart.

Krízsik, V. & Körmöczi, L. (2000). Spatial spreading of *Robinia pseudo-acacia* and *Populus alba* clones in sandy habitats. Tiscia 32: 3-8.

Legendre, P. & Fortin, M. J. (1989). Spatial pattern and ecological analysis. Vegetatio 80: 107-138.

Legendre, P. & Legendre, L. (1998). Numerical ecology (2nd Ed.). Elsevier Science B. V., Amsterdam.

Lippitt, C. D., Rogan, J., Toledano, J., Sangermano, F., Eastman, J. R., Mastro, V. & Sawyer, A. (2008). Incorporating anthropogenic variables into a species distribution model to map gypsy moth risk. Ecological Modelling 210: 339–350.

Lobo, J. M., Jiménez-Valverde, A. & Real R. (2008). AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17: 145-151.

Lodge, D. M. 1993. Biological invasions: Lessons for ecology. Tree 8: 133-137.

Marcot, B. G. (2006). Habitat modelling for biodiversity conservation. Northwestern naturalist 87: 56-65.

Masaka, K. & Yamada, K. (2009). Variation in germination character of *Robinia pseudoacacia* L. (Leguminosae) seeds at individual tree level. Journal of Forestry Research 14: 167-177.

McCune, B. & Grace, J. B. (2002). Analysis of Ecological Communities. Gleneden Beach, Oregon, USA.

Michener, D. C. (1988). The introduction of Black Locust (Robinia pseudoacacia L.) to Massachusetts, pp.52-57.

Ministry of Agriculture, Forestry and Food. URL: http://www.mkgp.gov.si/.

Mooney, H. A. & Drake, J. A. (1989). Biological invasions: a SCOPE program overview. Biological Invasions, John Wiley and Sons, New York, pp. 491-508.

Mingyang, L., Yunwei, J., Kumar, S. & Stohlgren, T. J. (2008). Modeling potential habitats for alien species *Dreissena polymorpha* in Continental USA. Acta Ecologica Sinica 28: 4253-4258.

Morimoto, J., Kominami, R. & Koike, T. (2009). Distribution and characteristics of the soil seed bank of the black locust (Robinia pseudoacacia) in a headwaterbasin in northern Japan. Landscape and Ecological Engineering 6: 193-199.

Mosher, E. S., Silander, J. A. J. & Latimer, A. M. (2009). The role of land-use history in major invasions by woody plant species in the northeastern North American landscape. Biological Invasions 11: 2317-2328.

Nascimbene, J. & Marini, L. (2010). Oak forest exploitation and black-locust invasion caused severe shifts in epiphytic lichen communities in Northeran Italy. Science of the Total Environment 408: 5506-5512.

Noble, I. R. (1989). Attributes of invaders and the invading process: Terrestrial and vascular plants. In: Biological Invasions: A Global Perspective, pp. 301-313. John Wiley & Sons ltd.

Novice: gospodarske, obertniške in narodne. (1858). 'O seji akacije'. Ljubljana.

Ogrin, M. (2009). Spreminjanje podnebja v Prekmurju po 2. Svetovni Vojni. In: Pomurje: Trajnostni Regionalni Razvoj ob Reki Muri. Zbornik povzetkov. Ljutomer, Murska Sobota.

Pauchard, A. & Alaback, P. B. (2004). Influence of elevation, land use, and landscape context on patterns of alien plant invasions along roadsides in protected areas of South-Central Chile. Conservation Biology 18: 238-248.

Paušič, A., Čarni, A. & Somodi, I. (2011, in press). Prostorka analiza in modeliranje vzorca pojavljanja mokrotnih travnikov (*Molinion caeruleae* Koch 1926) na Goričkem. Anali PAZU, Murska Sobota.

Pearson, R. G. & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12: 361-371.

Perko, D. & Adamič, M. O. (1998). Slovenija: Pokrajine in Ljudje. Založba Mladinska Knijga, Ljubljana.

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

Pimentel, D. (2005). Environmental consequences and economic costs of alien species. Invasive Plants: Ecological and Agricultural Aspects: 269-276.

Polechová, J. & Storch, D. (2008). Ecological Niche, in Encyclopedia of Ecology, Vol. 2: pp. 1088-1097, eds. Sven Erik Jørgensen and Brian D. Fath, Oxford: Elsevier.

Porter, W. P., Budaraju, S., Stewart, W. E. & Ramankutty, N. (2000). Calculating climate effects on birds and mammals: Impacts on biodiversity, conservation, population parameters, and global community structure. American Zoologist 40: 597-630.

Pyšek, P., Prach, K. & Mandák, B. (1998). Invasions of alien plants into habitats of Central European landscape: An historical pattern. Plant Invasions: Ecological Mechanisms and Human Responses, pp. 23-32.

R Development Core Team (2008). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org.

Radosevich, S. R., Stubbs, M. M. & Ghersa, C. M. (2003). Plant invasions: Process and patterns. Weed Science 51: 254-259.

Rédei, K. (2002). Management of black locust (*Robinia pseudoacacia* L.) stands in Hungary. Journal of Forestry Research 13: 260-264.

Rédei, K., Ostváth-Bujtás, Z. & Balla, I. (2001): Vegetative propagation methods for black locust (*Robinina pseudoacacia* L.) improvement. Hungarian Agricultural Research 10: 6-9.

Ricciardi, A. (2007). Are Modern biological invasions an unprecedented form of global change? Conservation Biology 21: 329-336.

Rice, S. K., Westerman, B. & Federici, R. (2004). Impacts of the exotic, nitrogen-fixing black locust (*Robinia pseudoacacia*) on nitrogen-cycling in a pine–oak ecosystem. Plant Ecology 174: 97–107.

Richardson, D. M., Pyšek, P., Rejmánek, M., Barbour, M. G., Panetta, D. F., & West, C. J. (2000). Naturalization and invasion of alien plants: concepts and definitions. Diversity and Distributions 6: 93-107.

Rudolf, S. & Brus, R. (2006). Razširjenost in invazivnost robinije (*Robinia pseudoacacia* L.) v severovzhodni Sloveniji. Slovenska strokovna revija za gozdarstvo 64.

Sax, D. F., Gaine, S. D. & Brow, J. H. (2002). Species invasions exceed extinctions on islands worldwide: A comparative study of plants and birds. The American Naturalist 160: 766–783.

Sebert-Cuvillier, E., Simon-Goyheneche, V., Paccaut, F., Chabrerie, O., Goubert, O. & Decocq, G. (2008). Spatial spread of an alien tree species in a heterogeneous forest landscape: a spatially realistic simulation model. Landscape Ecology 23: 787-801.

Soberón, J. & Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species distributional areas. Biodiversity Informatics 2: 1-10.

Somodi, I., Virágh, K., Székely, B. & Zimmermann, N. E. (2010). Changes in predictor influence with time and with vegetation type identity in a post-abandonment situation. Basic and Applied Ecology 11: 225-233.

Stohlgren, T. J., Bull, K. A., Otsuki, Y., Villa, C. A. & Lee, M. (1998). Riparian zones as havens for exotic plant species in the central grasslands. Plant Ecology 138: 113-125.

Swearingen, J. (2009). Weed US database of plants invading natural areas in the United States: Black Locust (*Robinia pseudoacacia*). http://www.invasive.org/weedus/subject.html?sub=3350. Consulted on March 20, 2011.

Thuiller, W., Richardson, D. M., Pysek, P., Midgley, G. F., Hughes, G. O. & Rouget, M. (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology 11: 2234-2250.

Torelli, N. (2002). Robinija (*Robinia pseudoacacia* L.) in njem les. Revija Les 54: 6-10.

Trethowan, P. D., Robertson, M. P. & McConnachie, A. J. (2011). Ecological niche modelling of an invasive alien plant and its potential biological control agents. South Africa Journal of Botany 77: 137-146.

Turner, M. G. (1987). Spatial simulation of landscape changes in Georgia: A comparison of 3 transition models. Landscape Ecology 1: 29-36.

Veenvliet, J. K., Veenvliet, P., Bačič, M., Frajman, B., Jogan, N. & Krajšek, S. S. (2009). Tujerodne vrste ubežnice z vrtov. Zavod Symbiosis, Grahovo.

Vitousek, P. M., D’Antonio, C. M., Loope, L. L., Rejmánek, M. & Westbrooks, R. (1996). Biological invasions as global environmental change. American Scientist 84: 468-478.

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

Walker, L. R. & Vitousek, P. M. (1991). An invader alters germination and growth of a native dominant tree in Hawaii. Ecology 72: 1449-1455.

Walter, J., Essl, F., Englisch, T. & Kiehn, M. (2005). Neophytes in Austria: Habitat preferences and ecological effects. In: Nentwig, W. et al. (Eds.): Biological Invasions – From Ecology to Control. NEOBIOTA 6: 13-25.

Wieseler, S. (1998). Black Locust. Plant Conservation Alliance. http://www.nps.gov/plants/alien/fact/rops1.htm. Consulted on March 20, 2011.

Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. (1998). Quantifying threats to imperiled species in the United States: Assessing the relative importance of habitat destruction, alien species, pollution, overexploitation and disease. Bioscience 48: 607–615.

Wood, S. N. (2006). Generalized Additive Models: An Introduction with R. Boca Raton, FL: Chapman & Hall/CRC.

Wraber, M. (1959). Fitosociološka opredelitev Žitkovskega, Kobiljanskega, Bukovniškega in Rediškega gozda v Prekmurju. 35 pages. Institut za Biologijo, ZRC-SAZU. Ljubljana.

Wraber, M. (1964). Gozdne vegetacijske enote in rastiščnogojitveni tipi na Goričkem v Prekmurju. 52 pages. Institut za Biologijo, ZRC-SAZU. Ljubljana.

Wraber, M. (1968). Gozdna vegetacija Ginjevca, Hraščice in Orlovščka v Pomurju. 26 pages. Institut za Biologijo, ZRC-SAZU. Ljubljana.

Zavod za Gozdove Slovenije. (2009). OE Murska Sobota.

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

Zhu, L., Sum, O. J., Sang, W., Li, Z. & Ma, K. (2007). Predicting the spatial distribution of an invasive species (*Eupatorium adenophorum*) in China. Landscape ecology 22: 1143-1154.

## Acknowledgements

First I would like to acknowledge doc. dr. Andraž Čarni, my mentor, for his encouragement and guidance throughout this master degree.

Many of my ideas regarding the methodology part of this study were formed in one year of consultation with dr. Imelda Somodi, from the Institute of Ecology and Botany from the Hungarian Academy of Sciences, to whom I'm especially thankful. I would also like to thank the Eötvös Loránd University, Budapest, Hungary who hosted me during my two weeks training.

The fieldwork conducted for this study was partly supported by the TransEcoNet project implemented through the CENTRAL EUROPE Programme and co-financed by the ERDF.

I would like to recognize the Laboratory for Environmental Sciences, especially to the head of the Laboratory Prof. dr. Urška Lavrenčič Štangar, for funding part of my studies.

I am grateful to the Scientific Research Centre of the Slovenian Academy of Sciences and Arts, especially to the Jovan Hadži Institute of Biology for their hosting, literature and software services.

Grateful acknowledgement is made to Prof. dr. Julia Ellis-Burnet for her English proofreading, which definitely improve the clarity and flow of this manuscript.

I would like to show my gratitude to Mr. Danilo Belak, forestry engineer from Zavod za Gozdove Slovenije (Slovenian forestry service) who was always willing to share with me information and data of the distribution of study species within the study area.

To the Surveying and Mapping Authority of the Republic of Slovenia to provide the data necessary for the analysis performed in this work.

My special thanks to my family for supporting me through the years and helped me to believe in myself.

To Matjaž who always stood by my side.