1-15-2013

Using Remote Sensing and GIS to Assess the Effects of Land Use/Cover Change and Geographic Variables on the Spread of Poisonous Invasive Giant Hogweed in Latvia

Simon Foteck Fonji

USING REMOTE SENSING AND GIS TO ASSESS THE EFFECTS OF LAND USE/COVER CHANGE AND GEOGRAPHIC VARIABLES ON THE SPREAD OF POISONOUS INVASIVE GIANT HOGWEED IN LATVIA.

by

Simon Foteck Fonji

A Dissertation

Submitted in Partial Fulfilment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Earth Sciences

The University of Memphis

May, 2013

Acknowledgements

This dissertation would not have been possible without the help of so many people around me. I will first of all, begin by thanking the almighty God for giving the strength, health and courage to carry out this work to the final phase.

A hearty thanks to all members of my committee: Dr. Esra Ozdenerol, Dr. Gregory N. Taff, Dr. Brian A. Waldron, Dr. Takuya Nakazato, Dr. Arleen A. Hill, and Dr. Anzhelika Antipova for their advice, suggestions and criticisms that greatly improved the quality of this research. Special thanks to Dr. Gregory Taff who despite having a new job in Norway continued to work with me and has greatly broadened my knowledge of remote sensing and statistics as tools used to conduct scientific research. You are a true mentor and I sincerely thank you for all your time during online conference meetings.

Thanks to all faculty members and students in the Department of Earth Sciences for their support, encouragement and shared ideas. Special thanks to Mike Larrivee and Rebecca Almond for all the work gathering GPS location points of Giant Hogweed during our field trip to Latvia in the summer of 2010. I am indebted to Agnese Priede who provided me with some GIS layers of Latvia and related web sites when I badly needed them. Many thanks to SAGE Lab. members (Cem Akkus and William DeVore) for GIS related problems and all the dinners and breakfast we had together.

I am very grateful for the contributions that made this work financially possible. This research was supported by the University of Memphis College of Arts and Sciences through the Travel Enrichment Fund, a University of Memphis Study Abroad Scholarship, and the Department of Earth Sciences at the University of Memphis. A special thanks to the Vidzeme University of Applied Sciences in Valmiera, Latvia for their assistance to support field work.

Thank you to my family and friends in Cameroon for all the prayers, encouragement and jokes that kept me strong. Special thanks to my sister Winifred and my cousin Wolfgang for all the necessary paper work. Thanks to the Musa family in Marietta Georgia for their hospitality and financial support. Uncle Joe you are a special guy, may God bless you.

Abstract

Fonji, Simon Foteck. Ph.D. The University of Memphis. May, 2013. Using Remote Sensing and GIS to assess the effects of land use/cover change and geographic variables on the spread of poisonous invasive Giant Hogweed in Latvia. Major Professor: Dr. Esra Ozdenerol.

Land-use and land-cover change (LULCC), especially those caused by human activities, is one of the most important components of global environmental change (Jessen, 2005). This dissertation analyzes the effects of geographic, biophysical, and demographic factors on LULCC and how LULCC and geographic variables influence the spread of invasive Giant Hogweed in northeastern Latvia. Data sets used in this study include: remote sensing images (Landsat Thematic Mapper acquired in 1992 and 2007), global positioning system (GPS data), census data, and data from public participation geographic information system (PPGIS). These data sets were processed and analyzed in a geographic information system (GIS). Six categories of land-cover were studied to determine land-cover change (LCC) and the relationship to population change between 1992 and 2007. Classification and analysis of the 1992 and 2007 Landsat images revealed that land-cover changing to forest is the most common type of change (17.1% of pixels) followed by changes to agriculture (8.6% of pixels) and the least was changes to urban/suburban (0.8% of pixels). Integration of the census data and land-cover classification revealed interesting patterns, for example, that population density is positively correlated with percent change in forest, agriculture and urban. Modeling the spread of Giant Hogweed was achieved using logistic regression and a novel cluster analysis approach. The logistic regression model was used to model the spread of Giant Hogweed using presence and pseudo-absence data of Giant Hogweed, while cluster analysis used only Giant Hogweed presence data. Both models were run using data from a series of GIS layers including topographic and land-use land-cover change (LULCC) information. The results from logistic regression and cluster analysis show that Giant Hogweed is likely to grow near roads, near rivers, in proximity to urban centers and in low elevation areas. Habitat suitability maps produced from both models indicate where Giant Hogweed is more likely to spread in the future and can serve as useful tools for policy makers and land managers to focus their efforts to manage weed invasions, and identify similar habitats where Giant Hogweed may occur in the future.

Table of contents

List of Tables

List of Figures

Chapter 1

Introduction

PhD. Dissertation in three parts.

Land cover and land use are often used interchangeably but the two have different meanings. Land cover describes the natural and anthropogenic features that can be observed on the Earth`s surface. Examples include deciduous forest, grassland, water and developed areas. Land use by contrast, describes activities, often associated with people that take place on the land and represent the current use of property. Examples include residential homes, shopping centers, row crops, tree nurseries and state parks. There are few landscapes remaining on Earth that have not been altered by humans. Humans have altered the landscape in such a way that has a profound effect on the natural environment. These anthropogenic influences on shifting patterns of land use are a primary component of many current environmental concerns as land use and land cover change is gaining recognition as a key driver of environmental change (Riebsame, et. al., 1994). Changes in land use and land cover are increasing rapidly, and can have adverse impacts and implications at local, regional and global scales.

To better understand the impact of land use land cover change (LULCC) on terrestrial ecosystems the factors that influence LULCC must be fully understood and the consequences. Growing human populations have increased pressure on the landscape as the demand for natural resources such as food, fuel and timber increases. Socio-economic factors often determine how land is used locally and regionally. LULCC has become an important component in strategies to manage natural resources and monitor environmental change. These are usually accomplished using remote sensing and GIS technologies. By utilizing remote sensing and implementing GIS mapping techniques, LULCC of designated areas can be monitored and mapped for specific research purposes and analysis.

The selection of the location for this research is based on the author's interest in LULCC and how it affects the spread of invasive species. To accomplish this, first I used LANDSAT TM image time series with 30m resolution to map and identify land cover changes in the study area and also to determine the drivers of change. Secondly, I want to investigate if there are any associations between LULCC and the occurrence of invasive Giant Hogweed. It has been observed that over the past decades, as the land cover changes so does the spread of invasive species has increased over time. In Latvia Giant Hogweed was introduce in the country in small farms to feed cattle but the plant has since then spread to other parts of the country. The plant is a successful invader for the following reasons: the plant mature to heights of 6 to 21 feet that makes it taller than any other herbaceous plant or grass in Latvia; purplish stalk can grow to 10 cm in diameter; leaves up to 1.5 meters across on lower 1/3 of stalk readily shade out competitors in the understory; compound umbel (flower) can reach 2 meters in diameter and produce up to 100,000 seeds with over 90% viability rate; huge taproot serves as nitrogen storage device and provides tremendous regenerative potential. In Latvia open areas such as cleared forest and abandoned agricultural areas are more susceptible to Giant Hogweed invasion.

This dissertation highlights the importance of human disturbance and geographic variables as factors influencing the spread of poisonous invasive Giant Hogweed in Latvia. Three complementary and interlinked methodologies were used to address the factors that promote Giant Hogweed spread taking into consideration the historical as well as the current distribution of the plant. In detailing how these factors influence Giant

Hogweed distribution and why Giant Hogweed thrive under certain conditions, this work aims to depict the interrelationship between the environment, humans and the spread of invasive species.

This research is divided into three parts and each part attempts to answer one of the three key questions:

- What are the local demographic factors associated with land use & land cover change in Latvia?
- Which land use/cover changes, geographic and socio-demographic variables are associated with higher occurrence of Giant Hogweed in Latvia?
- Can cluster analysis be used to accurately predict where Giant Hogweed will likely grow?

Part 1 of my dissertation examines the use of remote sensing data and Geographic Information Systems (GIS) in mapping land-cover change (LCC) in north eastern Latvia between 1992 and 2007 Unsupervised and supervised classification methods were used to classify LANDSAT satellite imagery, and post-classification change detection analysis was employed to determine changes in land cover. Subsequently demographic and geographic variables were used to determine the drivers of LCC in the study area.

Part 2 of my dissertation focuses on the effectiveness of logistic regression in modeling the spread of poisonous invasive Giant Hogweed in Latvia using LCC data and geographic variables as independent variables. The logistic regression identifies environmental factors that contribute to the proliferation of Giant Hogweed in the study region based on Giant Hogweed presence and absence data. A habitat suitability map was created using the regression equation and validated using test data. This paper is thus an attempt to begin to connect LCC and geographic variables to the occurrence of Giant Hogweed in Latvia.

Part 3 focuses on the use of cluster analysis (cluster of environmental variables as opposed to spatial cluster) to model the existing habitats of invasive Giant Hogweed in Latvia. Part 2 and Part 3 of my dissertation have the same goal but the two modeling techniques are different and produce somewhat different outcomes. One fundamental difference between the two methods is that while logistic regression uses both presence and absence data, cluster analysis uses presence data only. Using cluster analysis, four main clusters were identified and a habitat suitability map was produced based on inner-percentiles of cases within each cluster. This paper attempts to identify sites represented by the environmental clusters that have high or low Giant Hogweed presence based on a set of environmental variables.

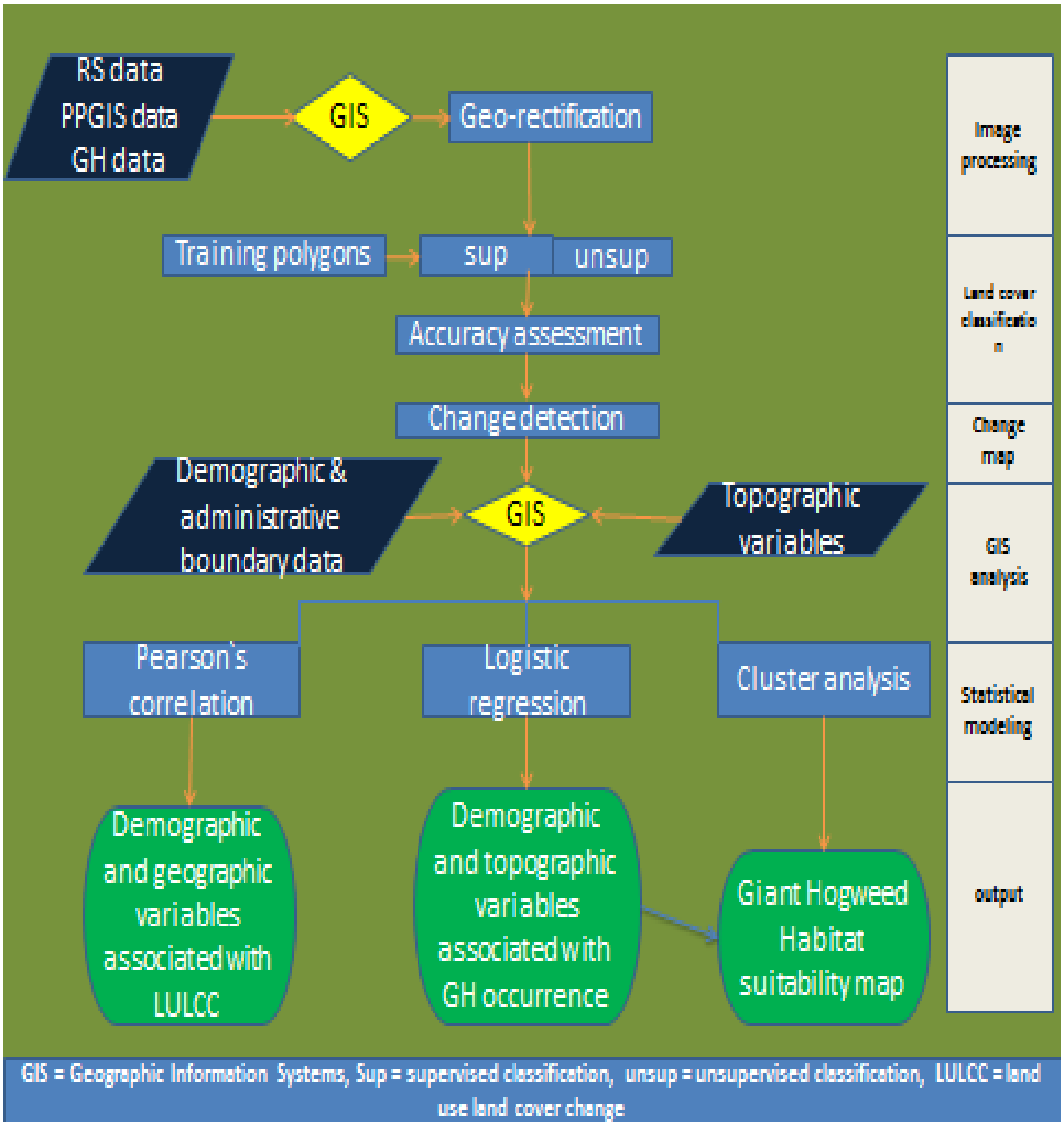

Fig 1 Project summary flow chart of land use land cover change map and Habitat suitability modeling of Giant Hogweed. Ground control points from field data collection was used to geo-rectify the LANDSAT TM images while ground truth data are used to classify the LANDSAT TM images. Logistic regression and cluster analysis were performed to identify important variables associated with Giant Hogweed occurrence and to produce a Giant Hogweed habitat suitability map. Arrows represent processes, boxes are mandatory developments.

References

Riebsame WE, Meyer WB, Turner II BL (1994) Modeling land use and cover as part of global environmental change. Climatic Change 28**:** 45-64.

Chapter 2

Using Satellite Data to Monitor Land use land cover Change in Northeastern Latvia.

Introduction

Land-use and land-cover are terms often used interchangeably but the two have different meanings. Land cover describes the natural and anthropogenic features that can be observed on the Earth's surface. Examples include deciduous forests, wetlands, developed/built areas, grasslands, and water. Land use, by contrast, describes activities, often associated with people that take place on the land and represent the current use of property. Examples include residential homes, shopping centers, tree nurseries, state parks, and reservoirs. LULCC especially those caused by human activities, is one of the most important components of global environmental change (Jensen 2005). According to Meyer and Turner (1992) LULCC is a hybrid category. Land use denotes the human employment of the land and is studied largely by social scientists. Land cover denotes the physical and biotic character of the land surface and is studied largely by natural scientists. Connecting the two are proximate sources of change: human activities that directly alter the physical environment. These activities reflect human goals that are shaped by underlying social driving forces. Proximate sources change the land cover, with further environmental consequences that may ultimately have feedback to affect land use. LULCC is local and place-specific, and collectively these changes add up to global environmental change. These changes in turn affect other components of the earth-atmosphere system, often with adverse consequences such as biodiversity loss, desertification, and climate change (Turner et al. 1990).

There are many ways to monitor or detect land cover change over time. In the past scientists used field data and aerial photographs to map LULCC. For large study areas, these methods can be very costly and time consuming. This is particularly true for remote regions, which are often inaccessible and thus not easy to obtain the needed data using traditional methods (Roberts et al. 2003; Cingolani et al. 2004). Remote sensing via satellite imagery is an excellent tool to study LULCC because images can cover large geographic extents, has a high temporal coverage and affords access to remote locations. Remote sensing is also used to investigate historical LULCC and also provide data (e.g. ground truth (Shao Yang et al. 2011; Asner and Warner 2003)) in areas that are inaccessible. The major disadvantages of remote sensing include: the inability of many sensors to obtain data and information through cloud cover, distinct phenomena can be confused if they look the same to the sensor; the resolution of the satellite imagery may be too coarse for detailed mapping and for distinguishing small contrasting areas and very high-resolution satellite imagery are very expensive. Despite these disadvantages, remotely sensed satellite data have been used to identify changes in a variety of aquatic and terrestrial environments including coastal, agriculture, forested, and urban areas (Berlanga and Ruiz 2002). LULCC researchers often use remotely sensed data to provide information on resource inventory and land use, and to identify, monitor and quantify changing patterns in the landscape.

Population change and distribution is a significant driver of land use land cover change in many regions of the world. With the emergence of GIS in the past two decades, census data have been merged with biophysical data to better understand the drivers of land use land cover change at local, regional and global scales. For example, the combination of satellite classification and census data has been used to assess quality of life (Lo and Faber 1997), predict favorable wolf habitat in northern Wisconsin (Mladenoff et al. 1995), assess the effect of population change on forest cover in Ghana between 1990 and 2010 (Codjoe 2004), understand the socioeconomic drivers of change in the Ecuadorian Amazon (Mena, Bilsborrow, and McClain 2006), understand change in agricultural activities in the Brazilian Amazon (Cardille and Foley 2003), comprehend relationships between land-cover and housing density in Wisconsin (Radeloff et al. 2000), and to study deforestation in the Brazilian Amazon (Wood and Skole 1998). This research integrates land cover change data (based on LANDSAT TM images from 1992 and 2007) and demographic data (from the Latvian demographic censuses and inter-census estimates based on vital statistics data for 1992 and 2007) at the level of rural municipalities (pagasti) and counties (rajoni) in a GIS to determine associations between LULCC and demographic factors (population density and population growth between 1992 and 2007). Considering the extent of current environmental problems such as land degradation, erosion, desertification, pollution, and invasive species, these research studies show clearly that GIS is an essential tool for the process of assessing and monitoring the impact of human activity and settlement patterns on spatial patterns and ecosystem dynamics, and for manipulating and displaying the information in ways that can be easily understood by those involved in studying the LCC overtime. Socio-demographic factors impact LULCC locally and regionally. The landscape of the study area, located in northeastern Latvia within in eastern Europe, is shaped by unique political, economic and socio-demographic factors influencing this region. Eastern Europe experienced a period of rapid and radical changes of its political, institutional, demographic, and socioeconomic structures after the fall of the Iron Curtain in 1989 and the breakdown of the Soviet Union in 1991 which triggered widespread land use change, most notably the abandonment of vast areas of cropland in Latvia and other countries in the region (Taff et al. 2010). Between the Second World War and the fall of the Soviet Union, private land ownership was forbidden in Latvia. After Latvia gained its independence from the Soviet Union in 1991, the government decided to reinstate lands to private owners. This period also saw a change from an agriculturally dominant economy to a capitalist system that drove much farm abandonment in favor of work in other economic sectors, generally located in large cities. From 2004 up till the first half of 2008 Latvia had the most rapidly developing economy throughout the European Union, with the GDP growth reaching 12.2% in 2006 (Eurostat 2010); then in 2009 the Latvian economy suffered a severe setback among the European Union Member States (GDP of -18%) forcing substantial outmigration to find employment. Land privatization and the increase in economic wealth during the 2000's drove substantial changes in land-use, particularly agricultural abandonment (Taff et al. 2010) and the recent development of urban sprawl in Latvia and many East European countries (Geyer 2011).

This research uses statistical analyses, remote sensing images, tomographic maps and census data to address the drivers of LCLUC. Two summer dates of LANDSAT Thematic Mapper images from 1992 and 2007 were processed, classified and analyzed, and an accuracy assessment was performed using ERDAS IMAGINE(Leica Geosystems) /ArcGIS (ESRI Inc.). Post-classification change detection was used to determine land use/cover changes between the two dates. Census data were then analyzed in conjunction with the land use/cover change results to understand associations between landscape changes and demographic and geographic variables. The specific research question this paper endeavors to answer is: how are local geographic and demographic factors associated with land use/cover change in Latvia in the time frame since its independence?

Landuse change categories analyzed in this study are agricultural abandonment and reforestation, new agricultural development, and urban/suburban development. While some research has shown that population increase rates do not always mirror urban/suburban development rates (Eetvelde and Antrop 2004), an increase in population is generally expected to accompany increases in urban and suburban development. Research has shown that agricultural abandonment is often, though not always, associated with outmigration from a region (Gellrich and Zimmermann 2007).

Materials and Methods

Study area

The study is carried out in Latvia which lies on the Baltic coast, in the northern part of Eastern Europe. Latvia is one of the three Baltic States situated on the east side of the Baltic Sea, the others being Estonia (to the north) and Lithuania (to the south). Latvia also borders Russia and Belarus to the east (fig. 1). The total land area of Latvia is 64.6 thousand km² and the terrain is mostly low plain, with the majority of the territory between 40-200 meters above sea level (Eberhards 1984). The climate is wet with moderate winters for this latitude. The average amount of precipitation is 600-650 mm annually; the vegetation period usually lasts for 180-200 days (Normunds 1993). The landscape is characterized by matured forests, secondary forest, meadows, farmland, abandoned farmlands, lakes, rivers, hills, plains, villages and dispersed rural homesteads (Bunkse 2000).

Latvia had the fastest growing economy in Europe from 2004 until the middle of 2008 (Eurostat, 2010). After this period, Latvia experienced a particularly severe economic setback along with the world economic crisis, which forced substantial outmigration for work opportunities. The increased out-migration, in addition to one of the lowest birth rates in the world, caused an increase in annual population loss (Zvidrins 1998). Most of Latvia’s export is from wood and agricultural products (Eurostat 2010). Therefore, maintaining the health of Latvia’s rural ecosystems is important to the economy of Latvia. During the summer of 2011, the author, along with a research team, collected field data of ground control points and ground truth of relevant land-cover types in the study area, which lies in the country's north-east, in the Vidzeme region, and parts of the Latgale and Zemgale regions (Fig.1). This region was chosen because of the wide variety of demographic changes in the region, including areas of out-migration and some of the few areas within Latvia experiencing in-migration, and it encompasses some of the fastest developing cities in the country, such as Valmiera, Sigulda, and Cesis.

Fig. 1 Study area - northeastern Latvia

Data

In this study two LANDSAT Thematic Mapper (TM) images (path 186, row 20) with a spatial resolution of 30x30 meters were acquired from the summers of 1992 and 2007. Summer images were chosen to best distinguish the spectral signatures of the different land cover types, and near-anniversary dates were chosen for consistency between the two time points. Cloud cover in the southwest of the image required creating a subset of the image without that portion (9% of image). Municipal boundaries and population data from the Latvian Central Statistical Bureau (obtained and adapted from Latvian Censuses and updated via estimates from vital statistics data) are also used in this study.

Geometric correction

Ground control points were collected during field work in the summer of 2011, to georectify the 2007 image. The 1992 image was then co-registered to the 2007 image in ERDAS IMAGINE 9.3. Both images were registered to a common Universal Traverse Mercator (UTM) projection. The 2007 image was geometrically corrected using the second-order polynomial method and re-sampled using the nearest neighbor technique to match the 1992 resolution. The nearest neighbor technique was used because it tranfers the original pixel values which is essential for post-classification change detection. One limitation of this method is that the edges of the image may appear jagged. The total RMS error for the 2007 image was approximately 2.5 meters, which is quite good for this study given that the RMS error should be at most half the size of the pixel (15m) (Campbell 2002). Both images were overlaid on top of each other and zoom to locate features (major highways) that are clearly visible on both images to make sure those features were directly on top of each other and displacement was minimal through visual inspection.

Land cover classification scheme

This study used per-pixel supervised classifications which group satellite image pixels with the same or similar spectral reflectance features into the same information categories (Campbell 2002). In addition to using relevant land-use and land-cover classes, all classes of interest must be carefully selected and defined to successfully classify remotely sensed data into land-cover (or land-use) information (Gong and Howarth 1992). Six information classes of interest to this study were chosen from the U.S. Geological Survey Land Use/Land Cover Classification System (Anderson et al. 1976; Vogelmann et al. 2001) based on dominant land-cover types in Latvia and the goal of this study, which was focused on identifying and analyzing the causes of land-use and land-cover changes associated with forest clearing, forest growth (including agricultural abandonment), and development. The six classes used were urban/suburban (built-up), agriculture, bare field/barren, forest, water, and wetlands. Bare soil was used as an additional class for classification of the LANDSAT images, but for land-use change analyses the agriculture and bare soil classes were merged into an agriculture land-use class, since bare soil in Latvia generally represents agriculture – fallow or recently harvested.

Maximum likelihood supervised classifications were performed in ERDAS IMAGINE 9.3 on the 1992 and 2007 LANDSAT images. For each class, 10 ground-truth polygons were digitized based on air photos and visual analysis of locations on Google Maps and the image itself. In order to insure proper alignment of the air photos and the LANDSAT image the transparency on the air photo was reduced to 50% and check with the base map to see which features on the map can also be seen on the air photos. Points such as rail roads, bends in roads, bends in rivers or river junctions, road intersection, buildings were checked especially in the corners of the air photos. There was no need for rotation.

In accordance with Jensen (2005), each polygon contained at least 50 pixels except in a few cases where a high proportion of land cover patches in the study area contained fewer than 50 pixels (for the suburban and water classes). To improve classification, training polygons with confusing spectral signatures were discarded and new ones created based on a visual analysis of the locations on Google Maps and on the image itself, and these were added to the existing samples and the maximum likelihood algorithm was run again. Pixels throughout the classified image were visually compared with the raw image and Google Maps to determine general classification accuracy. This process was repeated until the observable errors in the classification were negligible (similar spectral signatures on both the classified and raw image). Fig. 2a and 2b shows the final output of the supervised classification, which consists of two classified maps of northeast Latvia in 1992 and 2007.

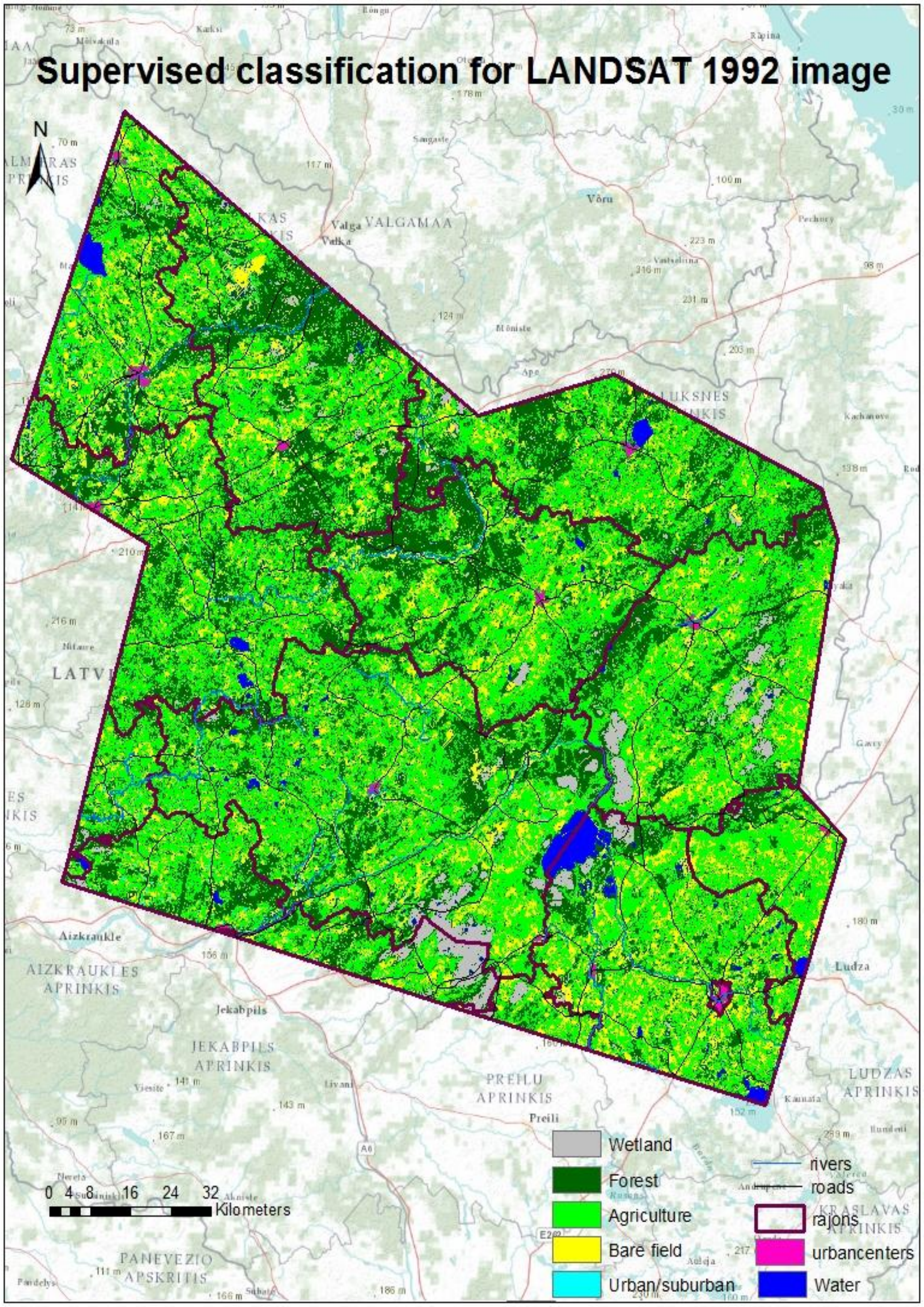

Fig. 2a Map of supervised classification of 1992

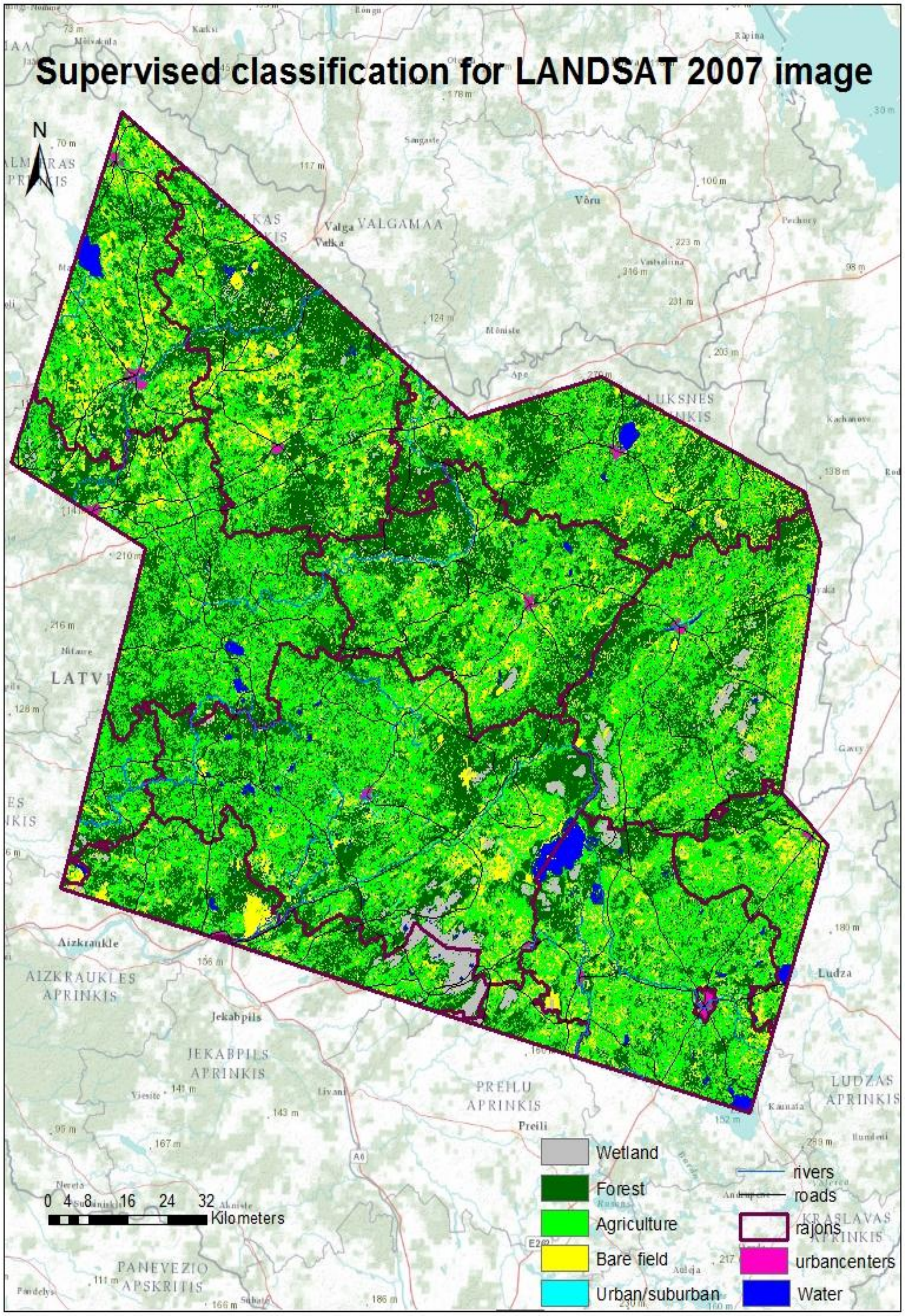

Fig. 2b Map of supervised classification 2007

Results

Accuracy assessment

An accuracy assessment was performed for the 2007 image, because of good available ground truth at this date. For the 2007 image, three hundred pixels were selected using stratify random sampling (50 pixels from each class). Stratify random sampling was used because it preserves the major randomness and eliminates the probability for an unbalance distribution of points within the categories. These selected pixels were then checked for accuracy using fieldwork ground truth data (2009 and 2010) when available, aerial photographs (2007) and Google Earth maps (2007). Out of the 300 Pixels, those that fell on the boundary between two classes and could not be classified as belonging to either class were eliminated. In total 15 pixels were eliminated and 285 pixels were checked for accuracy assessment. Three summaries are standardly reported for the accuracy assessment: the error matrix, the overall accuracy and the Kappa coefficient (Congalton 1991). Error matrices illustrate a quantitative comparison of the relationship between the classified maps and the reference data which may include field observations, aerial photographs or high resolution satellite images. The overall accuracy for the 2007 classified map based on the supervised classification was 89.12% which is considered good, and it is above the limit set by USGS guideline of 85% (USGS, 1996). User`s and producer`s accuracies were calculated to summarize classification accuracies for each class. Both the user`s and producer`s accuracies were very high in water (100% for all testing data). This is common and probably due to the fact that the spectral signature for water is quite different from the other classes. The user`s accuracy was also high for wetland (100%), forest (84.21%), bare field (92.11%), and urban/suburban (97.56%). The user`s accuracy was moderate for agriculture (71.43%) and this can be attributed to the commission errors from urban/suburban and bare fields. The producer`s accuracy was high in wetlands (86%), forest (96%), agriculture (90%), and urban/suburban (83.33%). The producer`s accuracy was moderate in bare field (77.78%). The bare field class contains errors of omission from agriculture and forest. Because the overall accuracy assessment tends to overestimate the actual performance, a more useful representation of performance is the kappa coefficient (Cohen 1960). The Kappa statistic for the supervised 2007 image is 0.869 which means that 86.9% of the classification is better than a random classification. This is considered good because a Kappa value above 80% is considered to have a strong agreement (Ramita et al. 2009). Table 1 reports the results for the accuracy assessment for the supervised classification of the 2007 image.

Table 1 Classification accuracy assessment using error matrix

Accuracy Assessment
for Supervised classification

		Class types determined from reference source							
		Water	Wetland	Forest	Agriculture	Bare field	urban/suburban	totals	user`s Accuracy
class types determined from classified map	Water	50	0	0	0	0	0	50	100.00%
	Wetland	0	36	0	0	0	0	36	100.00%
	Forest	0	0	48	3	5	1	57	84.21%
	Agric	0	5	2	45	5	6	63	71.43%
	Bare field	0	1	0	1	35	1	38	92.11%
	urban/ suburban	0	0	0	1	0	40	41	97.56%
	Totals	50	42	50	50	45	48	285	
Producer`s accuracy		100%	86%	96.00%	90.00%	77.78%	83.33%	Total Accuracy:	89.12%

Overall Kappa statistics = 0.869

Land-use change analysis

Bare soil and agriculture were recoded as one class because most of the bare soil land-cover class generally represented land-uses that were either pastures or agricultural fields without crops at that moment (e.g., recently harvested or fallow). A total of five classes were produced for each of the two images (water, wetlands, forest, agriculture, and urban/suburban). The 1992 image and the 2007 image were compared in terms of the total area of each land cover category. As seen in fig. 3, the classes that have increased in area include wetland, forest, and urban, while water, and agriculture have decreased over the same time period. For example 598 km^2 of area was wetland in 1992, and in 2007 it increased to 705 km^2. Out of the 274 km^2 area that was water in 1992, only 228 km^2 remained water in 2007. Water decrease can be attributed to water bodies drying up over years or within a season. Agriculture areas have decreased primarily due to farm abandonment and development (construction of buildings, roads, and houses). The lower percentage of agriculture can be attributed to farm abandonment that started primarily in 1991 after Latvia gained independence from the former Soviet Union and the agricultural sector became less profitable due to the breakup of farms to smaller plots (through land restitution) and the shift to a capitalist economy. Increase in forest may also be due to farm abandonment that resulted in the conversion of many agricultural fields into young forest. The persistence beyond the early post-Soviet transition years of increase in forest cover follows the forest transition theory (Rudel et al. 2005), which hypothesizes that forest cover decreases with early development levels of a society, but then increases with an even higher level of development. This period also saw a population shift within Latvia to bigger cities in search of jobs which led to a recent development of urban sprawl (increased buildings, and roads, at the urban periphery and beyond). Development can be seen throughout the study area, especially in cities and suburban areas of Valmiera, Cesis, Sigulda, and Madona.

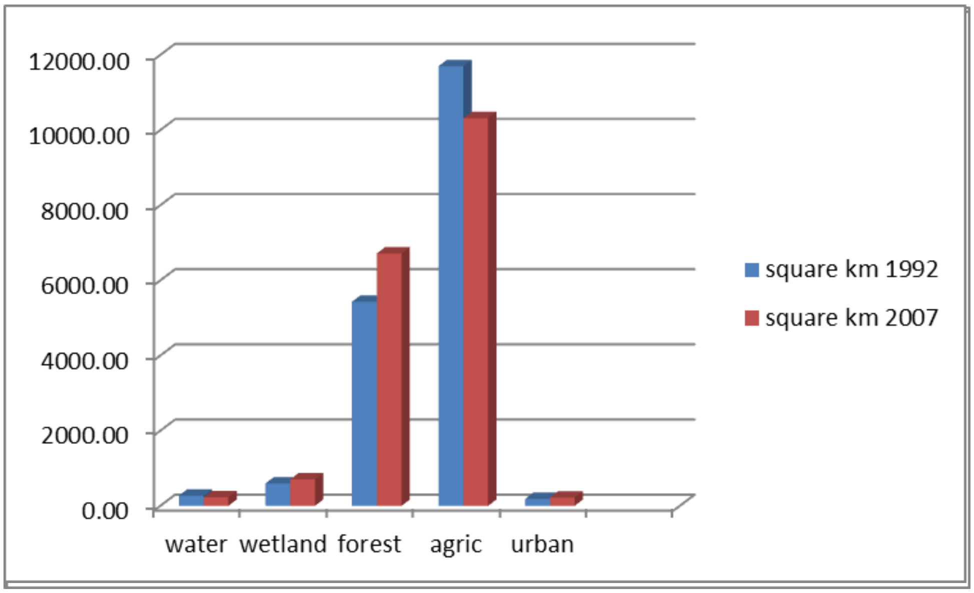

Fig. 3 Land-cover classes and area represented by each class in square kilometers for 1992 and 2007

A pixel-level "from-to" change analysis was then run with five classes and the result was a change map with twenty five classes. In order to reduce the number of classes in the change map, "no-change" classes and "change" classes that were considered unimportant (change to water or wetland because Giant Hogweed does not grow in water or wetlands) to the study were classified as irrelevant (with a value of 0), all the classes that *changed to* forest were classified as 1, all classes that *changed to* agriculture were classified as 2, and all classes that *changed to* urban were classified as 3 (fig. 4).

Among change classes, land cover types that changed to forest were most common (17.1%), followed by classes that changed to agriculture (8.6%), and finally classes that changed to urban (0.8%). The "No-change" class (71.6%) combined with changes considered unimportant (2%) for this study comprised 73.6% of the pixels as shown in table 2.

Table 2 Detail LULCC classes, 1= water, 2 = wetland, 3 = forest, 4 = agriculture, 5 = urban. Examples of LULCC: 12 = change from water to wetland, 34 = forest to agriculture, 43 = agriculture to forest, other changes = change from any land cover type to water or wetland.

LULCC class	Number of pixels	Area in square km	Area in percent
12	442	0.3978	0.00%
13	43914	39.5226	0.27%
14	16567	14.9103	0.10%
15	6	0.0054	0.00%
21	1087	0.9783	0.01%
23	167351	150.6159	1.05%
24	70071	63.0639	0.44%
25	544	0.4896	0.00%
31	5111	4.5999	0.03%
32	120252	108.2268	0.75%
34	1193399	1074.0591	7.47%
35	13396	12.0564	0.08%
41	5623	5.0607	0.04%
42	175590	158.031	1.10%
43	2509379	2258.4411	15.71%
45	111332	100.1988	0.70%
51	38	0.0342	0.00%
52	3441	3.0969	0.02%
53	5770	5.193	0.04%
54	85403	76.8627	0.53%
No change	11441425	10297.28	71.64%
Other changes	311584	280.45	1.95%
Total number of pixels	15970141	14373.1269	

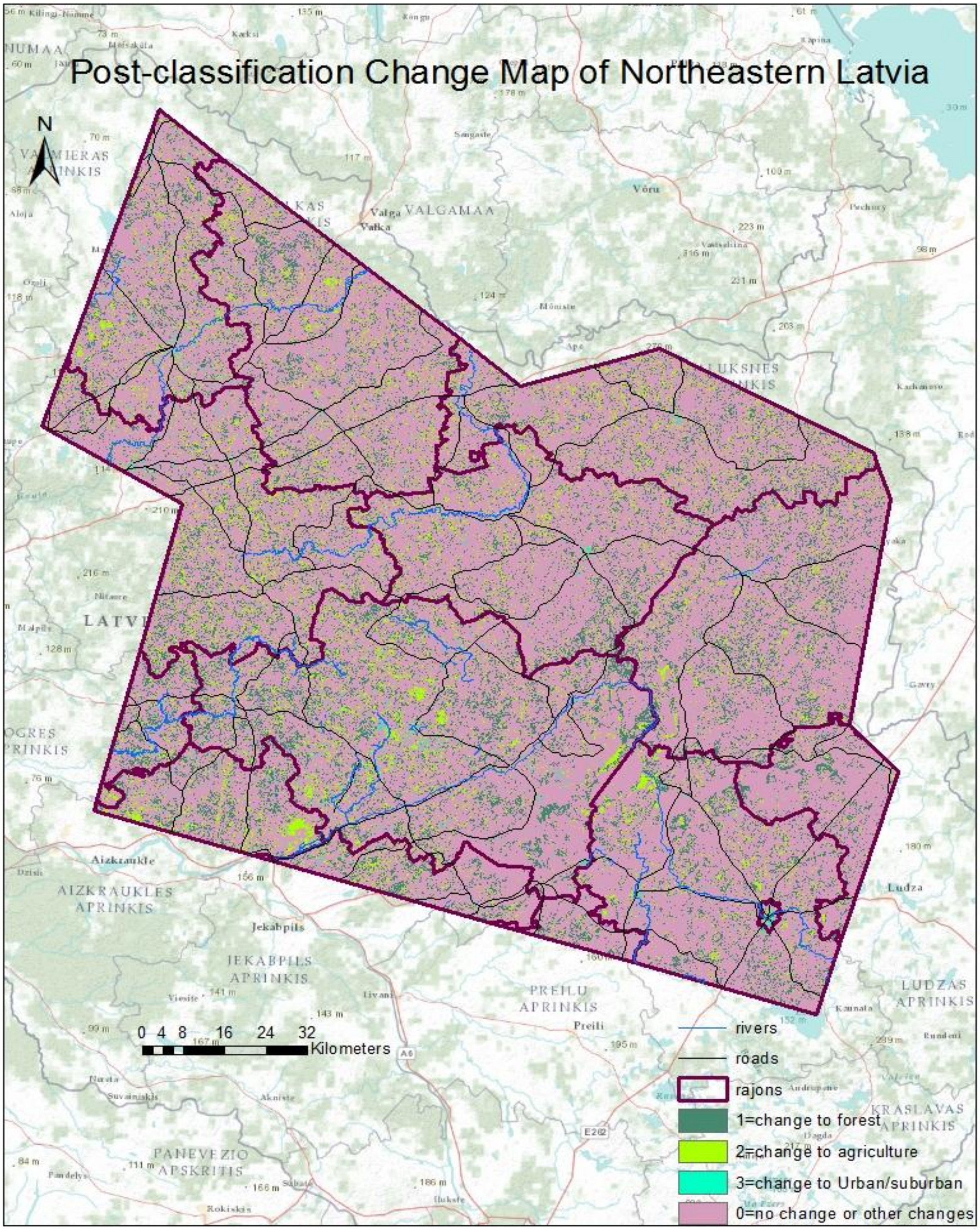

Fig. 4 Post-classification "change-to" map

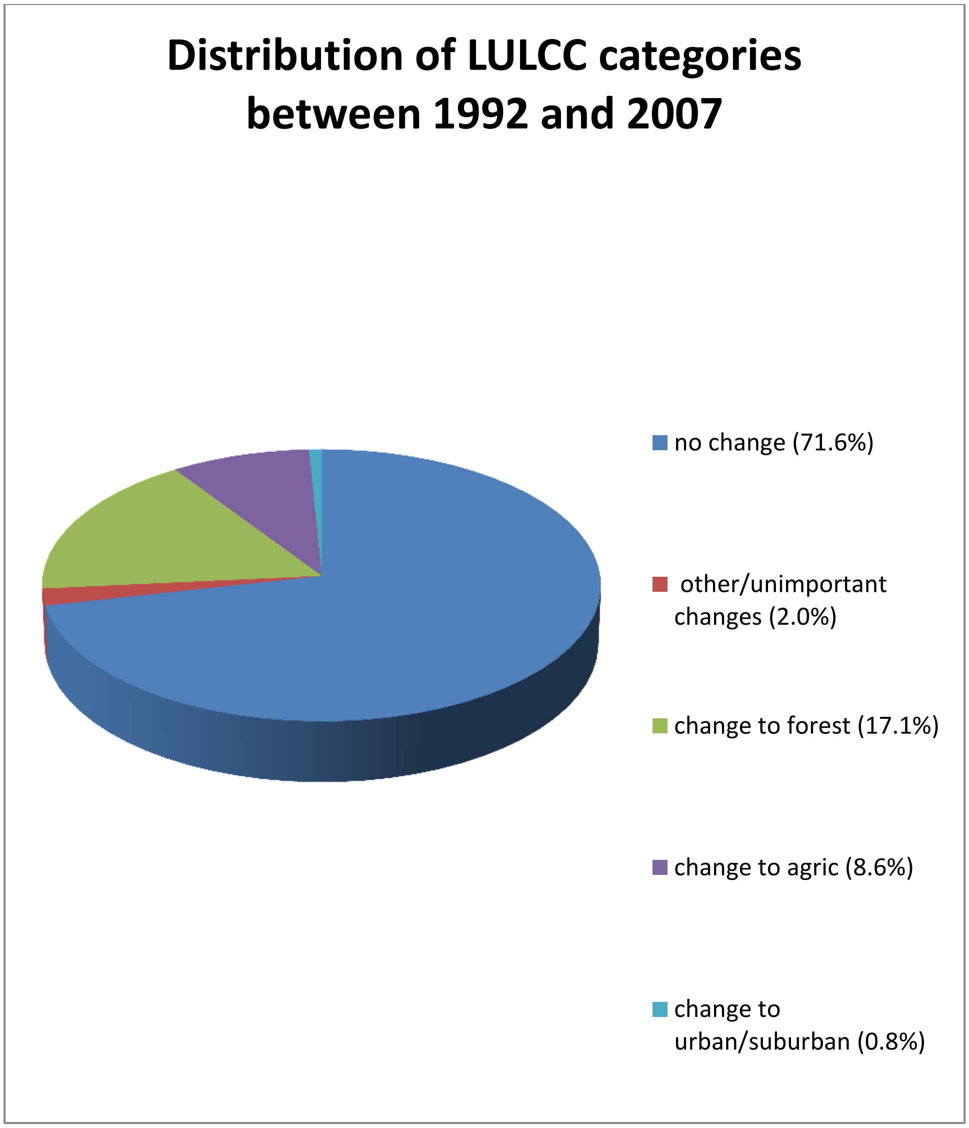

Fig. 5 Distribution of LULCC categories between 1992 and 2007 of Northeastern Latvia

Finally, the land use land cover change map was overlaid with municipal boundary data at the level of pagasti (municipality) and rajoni (similar to US counties) as shown in fig. 6 and 7.

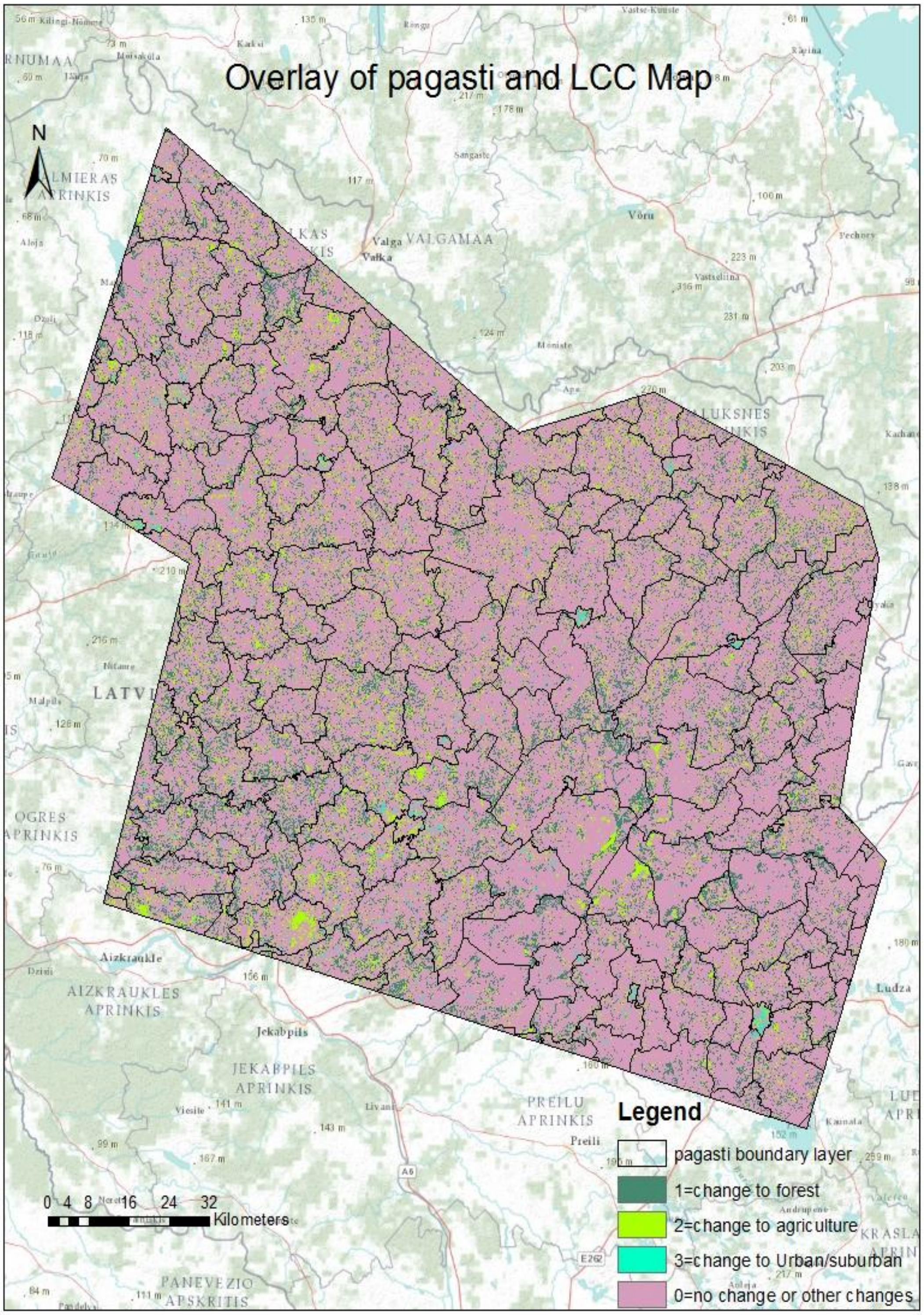

Fig. 6 Multilayer mapping of pagasti boundaries of Northeastern Latvia with land cover change map of LANDSAT TM image.

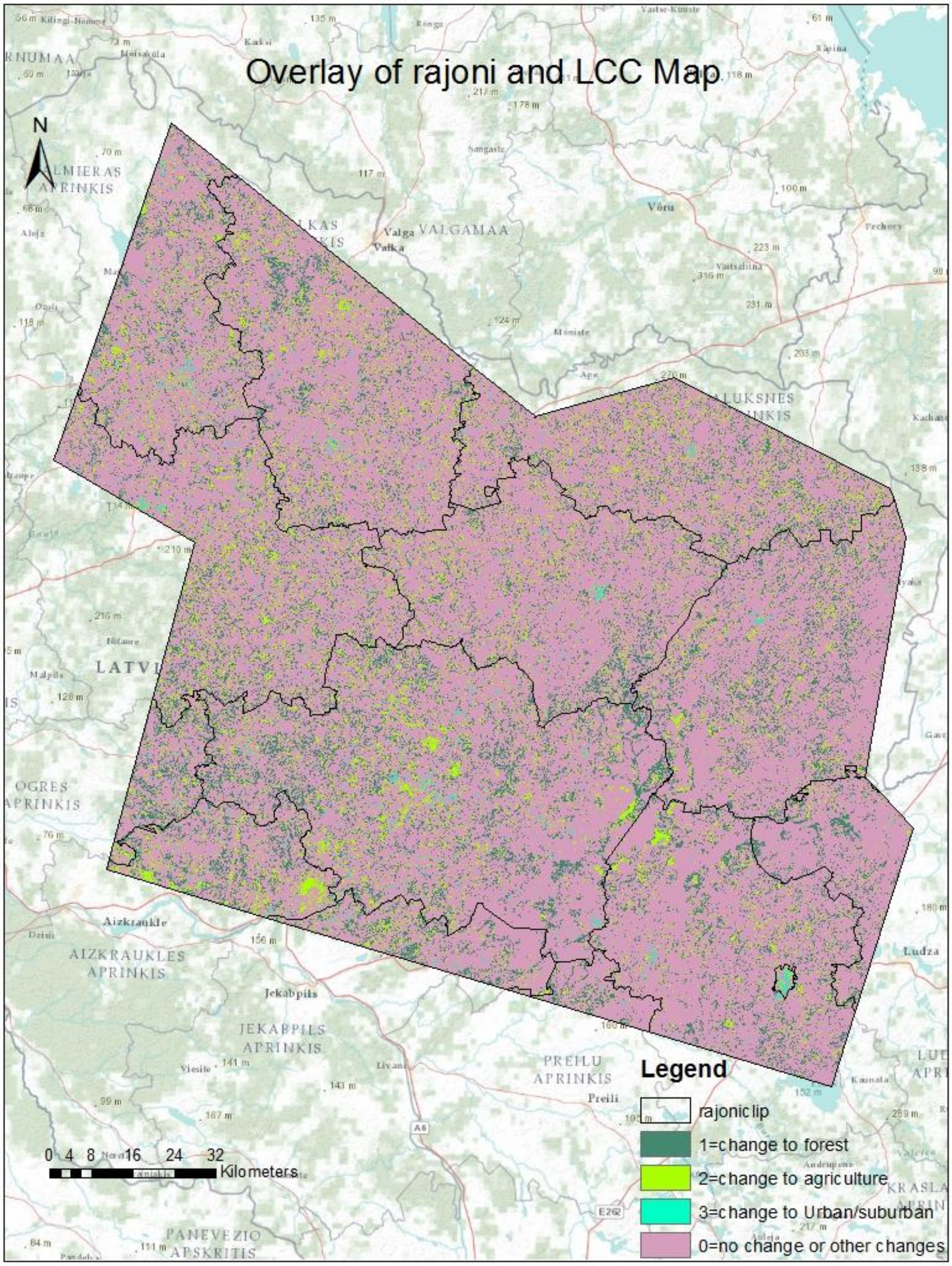

Fig. 7 Multilayer mapping of the rajoni boundaries of Northeastern Latvia with land cover change map of LANDSAT TM images.

Land use land cover classes were then extracted and aggregated at these two spatial units of analysis, namely rural pagasti (excluding big cities) and rajoni (including all major cities). The percent of each LULCC class was calculated in each pagasti and rajoni. Then a series of correlations were performed to establish the correlation between 1) socio-demographic factors (population density in 2008, percent population growth rate between 1993 and 2008,) and 2) geographic and topographic factors (elevation, % pagasti/rajoni area within given buffer distances from urban centers, % of protected areas within pagasti/rajoni and proximity to Riga) and the proportion of each LULCC class. Multiple buffer distances within 0.5 km -15km of percent pagasti and rajoni areas were created from roads and used to run the correlation. Buffer distances that were insignificant to all three LULCC types were eliminated. The results are shown in Table 2. The results show that, at the pagasti level, percent of pagasti area within 1km from roads and range in elevation are positively correlated with all three types of land cover change variables (i.e. percent change to agriculture, forest and urban/suburban). Therefore, it seems that the more number of roads a pagasti has (e.g. roads) the more rural development occurs (Creightney 1993). According to Chomitz and Gray (1996) rural roads not only promote economic development but they also facilitate deforestation. The significance with range in elevation is interesting; as a low range in elevation has been found in other studies to be associated with land-use changes, mostly occurring in coastal areas and in areas having low slope values (Selcuk 2008). In this study site, the higher ranges in elevation tend to follow the Gauja River Valley, along which fall the towns of Sigulda, Cesis and Valmiera; in addition, the areas near the slopes of this valley are popular tourist destinations. These two drivers likely promote development, reforestation (especially within Gauja National Park), and apparently agriculture as well (perhaps due to proximity to these rural towns). The positive relationship between range in elevation and the other types of land-use change may seem surprising (since development is not likely to occur at high elevations or on high slopes). But the range in elevation in the study area varies between 0-270m and most of the highest points are located near big cities and close to the Gauja River. Development in such areas may reflect scenic views. A similar study that revealed a positive correlation between range in elevation and land use change was reported by Turner et al. (1996). The percent of rajoni within 10km of an urban area was calculated, and found to be significantly negatively correlated with the change to forest. This means rajoni with more areas within 10km from urban centers experienced less changes to forest and is likely due to the higher value of lands near urban centers, both for agriculture and conversion to urban cover; therefore, less agricultural land was converted to forest in these areas. Percent population growth rate was not significantly correlated with percent change to forest and agriculture but it was positively correlated with percent change to urban at the level of pagasti. At the rajoni scale, percent population growth rate was not significantly correlated with any of the land cover variables. Percent of pagasti areas within 10km from urban centers and percent population growth rates were highly correlated with a change to urban/suburban development. These results are as expected: development occurs to accommodate population growth near cities. Also, the more increase in people a pagsati has the more urban development will occur to accommodate them. Factors that significantly negatively correlated with some land-use changes at the pagasti scale were population density in 2008 (decrease in population density in 2008) and proximity to Riga. New agricultural activities and forestry activities were occurring in pagasti with low population density. At the rajoni scale, there was a significant positive correlation between population density and change to forest, which is surprising, as it shows agricultural abandonment in high population areas; this may be due to the fact that most of the highest population rajoni have substantial area inside Gauja National Park, where forest regrowth was promoted through government policy (Taff 2005). Proximity to Riga was found to be negatively correlated with all land-cover change variables, and significant with some of them - on both the rajoni and the pagasti scales. Increased distance to Riga was associated with decreases in land-cover changes. Agriculture, forestry, and urbanization are happening more near Riga than far away. The result regarding urbanization is as expected, and the results regarding increased forestry and agriculture nearer to Riga may have to do with distance to market (cost of carrying goods to Riga) or because people are engaging in farming/forestry activities part time, while they or family members have other employment in Riga. Percent protected areas within pagasti/rajoni were not significantly correlated with any of the land cover variables at both the pagasti and rajoni scales. Also, range in elevation was not found to be significantly correlated with any of the land cover variables at the rajoni scale.

Table 3 Correlation of socio-demographic and geographic variables and LULCC at the level of pagasti and rajoni (*) significant correlation (p<0.05), (**) significant correlation (p<0.01).

	Socio demographic/ geographic factors	**% change to Agriculture**	**% Change to Forest**	**% Change to Urban/suburban**
Pagasti	% pagasti area within 10km of urban centers P-values	-0.029 0.72	-0.082 0.302	.175* 0.027
	% pagasti area within 1km from roads P-values	.213** 0.007	.162* 0.041	.209** 0.008
	Population density 2008 P-values	-.160* 0.043	-.236** 0.003	-0.068 0.394
	% Population growth 1993-2008 P-values	0.054 0.5	-0.076 0.338	.215** 0.007
	% Protected area within pagasti P-values	0.144 0.07	0.096 0.228	.180* 0.023
	Proximity to Riga in Km P-values	-.286** 0	-0.086 0.28	-.550** 0
	range in elevation P-values	.360** 0	.280** 0	.304** 0
	% rajoni area within 10km of urban centers	-0.289	-.533*	-0.078

	P-values	0.316	0.05	0.791
Rajoni	% rajoni area within 1km from roads P-values	-0.297 0.303	-0.34 0.234	-0.138 0.639
	% population growth 1993-2008 P-values	0.318 0.268	0.472 0.088	0.503 0.066
	Population density 2008 P-values	-0.032 0.913	.645* 0.013	0.526 0.053
	% Protected areas within rajoni P-values	0.002 0.994	-0.275 0.341	0.247 0.395
	proximity to Riga P-values	-0.523 0.055	-0.452 0.104	-.727** 0.003
	Range in elevation P-values	-0.181 0.535	0.043 0.883	0.231 0.427

Conclusion

Land-use and land-cover change monitoring in northeastern Latvia was achieved using post-classification change detection. The results demonstrate that remote sensing can be used to assess, monitor and quantify land-use and land-cover change in large areas where traditional methods (such as field observation) may not be possible. The results from this study revealed changes in some landscape patterns in northeastern Latvia in the post-Soviet period. The most significant land-cover change experienced in the study area was increase in forest cover. The percent of forest cover loss between 1992 and 2007 was 8.3% and percent of forest cover gain over the same time period was 17.1%. Out of the 17.1%, 15.71% resulted from the conversion of agriculture to forest cover. Much of this change can be attributed to reforestation (table 2) resulting from agricultural abandonment, as is common throughout the former Soviet Union since 1991 (Taff et al. 2010). Reforestation on abandoned agricultural lands was most common in low population density pagasti e.g. 70.2% of land cover that changed to forest occurred in pagasti that had less than ten people per square kilometers. The increase in urban/suburban cover occurred mostly around big cities and in high population density pagasti. Between 1992 and 2007, the percent of agriculture loss was 17.5% and percent of agriculture gain was 8.6%.

Among the socio-demographic and geographic variables used to explain land cover change in this region, the most contributing factors with positive correlation to all three land cover change variables were percent of pagasti/rajoni area within 1km from roads, and range in elevation. Proximity to Riga had a negative correlation with all the land cover change variables at both the pagasti and rajoni scale. Some of the factors that had a significant correlation at the pagasti scale (e.g. elevation) had no significant correlation at the rajoni scale. In general, based on comparison between the pagasti and rajoni correlations, it appears that the rajoni scale is too broad to catch the appropriate processes with regards to some of the demographic and geographic variables.

Latvia has been experiencing serious depopulation and low fertility levels (Zvidrins 1998; European Environment agency 2010) since independence in 1991. The demographic statistics show that most of the rural pagasti (147 out of 159) have experience a decrease in their population between 1993 and 2007. This research shows that regions of rural depopulation in Latvia are associated with substantial agricultural abandonment. While an increase in forest cover may increase the country`s capacity to sequester carbon, provide clean air, prevent erosion, and address some other environmental issues, it is worthwhile to note that decrease in agricultural land in this region can lead to decrease in biodiversity, loss in cultural landscape and some tourist opportunities (Taff 2005), and loss in food production in the region, which can lead to reliance on foreign imports and a trade imbalance. The food production index (covers food crops that are considered edible and that contain nutrients) in Latvia has dropped from 222.00 in 1992 to 138.00 in 2009 (World Bank Indicators 2010). These results on patterns of change in agriculture and associated variables can help policy makers to address key relevant variables associated with loss of agriculture, which can help them address food security, cultural landscape maintenance, and employment issues – the rate of unemployment in Latvia in 1992 was 3.178% and in 2010 was 18.965% (International Monetary Fund 2011). Considering the three land cover types in this study, the lowest percentage of land cover (0.8%) was converted to urban/suburban development, and most of these changes took place near cities. Results of this study therefore suggest the need to monitor increases in urban/suburban cover near cities, and to plan to maintain/develop green spaces. The general land-use change trends found in this study, and the variables found to be associated with these land-use changes, can be useful to policy makers beyond the study site in the northeast to all of Latvia regarding land-use planning based on geographic and population characteristics to address issues of land-use, economic development, and food security.

In this study significant correlation was found between demographic and geographic variables and LULCC data which implies demographic and geographic conditions played an important role in local landscape change at both the pagasti and rajoni scales. Future studies should research the association of socio-demographic and geographic variables at the city scale and incorporate more economic variables such as poverty and unemployment.

Acknowledgements

This research was supported by the University of Memphis College of Arts and Sciences through the Travel Enrichment Fund, a University of Memphis Study Abroad Scholarship, and the Department of Earth Sciences at the University of Memphis. A special thanks to the Vidzeme University of Applied Sciences in Valmiera, Latvia for their assistance to support field work.

References

Anderson JR, Hardy E, Roach J, Witmer R (1976) A Land Use and Land Cover Classification System for use with remote sensor data. Washington DC: U.S. Geological Survey Profession Paper 964, p 28

Berlanga CA, Ruiz LA (2002) Land use mapping and change detection in the coastal zone of Northwest Mexico using remote sensing techniques. Journal of Coastal Research, 18(3): 514-522

Bunkse EV (2000) Reality of Rural Landscape Symbolism in the Formation of a Post-Soviet, Postmodern Latvian Identity. Norsk Geografisk Tidsskrift, Volume 53, pp. 121 – 138

Campbell JB (2002) Introduction to remote sensing (3rd ed.). The Guilford Press. ISBN 1-57230-640-8

Cardille JA, Foley JA (2003) Agricultural land-use change in Brazilian Amazonia between 1980 and 1995: Evidence from integrated satellite and census data. Remote Sensing of Environment 87(4): 551-562

Chomitz K, Gray D (1996) Roads, Land Use and Deforestation: A Spatial Model Applied to Belize, World Bank Economic Review, 10(3):487-512.

Cingolani A, Renison D, Zak M, Cabido M (2004) Mapping vegetation in a heterogeneous mountain rangeland using LANDSAT data: an alternative method to define and classify land-cover units. Remote Sensing of Environment, 92(1):84-97

Codjoe SN (2004) Population and land use/cover dynamics in the Volta River Basin of Ghana. Ecology and Development Series, No. 15. Gottingen: Cuvillier Verlag

Cohen J (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37-46

Cohen WB, Goward SN (2004) LANDSAT's role in ecological applications of remote sensing. BioScience 54:535-545

Congalton GR (1991) A review of assessing the accuracy of classification of remote sensed data. Remote Sensing and Environment 7:35-46

Creightney, Cavelle D (1993) Transport and Economic Performance: A Survey of Developing Countries. World Bank Technical Paper 232. Washington, D.C.

Eberhards G (1984) Reljefs (The relief). In: Latvijas enciklopedija (Encyclopedia of Latvia), Vol. 5, part 2, GER, Riga, pp. 34-41

Eetvelde V, Antrop M (2004) Analyzing structural and functional changes of traditional landscapes-two examples from Southern France. Landscape and Urban Planning 67(1-4):79-95

European Environment Agency (2010) State and outlook, country assessment - Latvia http://www.eea.europa.eu/soer/countries/lv/soertopic_view?topic=land.Accessed 12 April 2012

Eurostat (2010) Economy of Latvia. http://balticexport.com/?article=latvijas-ekonomika&lang=en Accessed 12 April 2012

Geyer HS (2011) International Handbook of Urban Policy and Issues in the Developed World. UK: Edward Elgar Publishing Limited

Gellrich M, Zimmermann NE (2007) Investigating the regional-scale pattern of agricultural land abandonment in the Swiss mountains: A spatial statistical modeling approach. Landscape and Urban Planning 75(1): 65-76

Gong P, Howarth P (1992) Frequency-based Contextual Classification and Gray-level Vector Reduction for Land-use Identification, Photogrammetric Engineering and Remote Sensing 58(4):423-437

Gregory P. Asner, Amanda S. Warner 2003. Canopy shadow in IKONOS satellite observations of tropical forests and savannas. Remote Sensing of Environment 87, 521 – 533

International Monetary Fund (2011) World Economic Outlook, http://www.indexmundi.com/latvia/unemployment_rate.html. Accessed 12 April 2012

Jensen JR (2005) Introductory Digital Image Processing: A Remote Sensing Perspective. New Jersey: Pearson Education, Inc.

Lo CP, Faber BJ (1997) Integration of LANDSAT Thematic Mapper and census data for quality of life assessment. Remote Sensing Environ. 62(2): 143-157

Mena CF, Bilsborrow RE, McClain M (2006) Socioeconomic Drivers of Deforestation in the Northern Ecuadorian Amazon. Environmental Management, 37(6): 802-815

Meyer BW, and Turner B.L (1992). Human population growth and global land-use/cover change. Annu. Rev. Ecol. Syst. 23, 39-61

Mladenoff DJ, Sickley TA, Haight RG, Wydeven AP (1995) A regional landscape analysis and prediction of favorable gray wolf habitat in the northern great lakes region Conserv. Biol. 9(2):279-294

Normunds P (1993) Black Alder Swamps on Forested Peatlands in Latvia Folia Geobotanica & Phytotaxonomica, Vol. 28, No. 3, pp. 261-277 Published by: Springer

Radeloff VC, Hagen AE, Voss PR, Field DR, Mladenoff DJ (2000) Exploring the spatial relationship between census and land-cover data. Society and Natural Resources 13 (6) 599–609

Ramita M, Inakwu OA, Tiho A (2009) Improving the accuracy of land-use and land-cover classification of LANDSAT data using post-classification enhancement. Remote Sensing, ISSN 2072-4292. www.mdpi.com/journal/remotesensing

Roberts DA, Keller M, Vianei S (2003) Studies of land-cover, land-cover, and biophysical properties of vegetation in the large scale biosphere atmosphere experiment in Amazonia. Remote Sensing of Environment, 87: 377-388

Rudel TK, Coomes OT, Moran E, Achard F, Angelsen A, Xu J, Lambin E (2005) Forest transitions: towards a global understanding of land use change. Glob Env Change A 15(1):23-31

Selcuk R (2008) Analyzing Land Use/Land Cover Changes Using Remote Sensing and GIS in Rize, Noth-East Turkey. Sensors. ISSN 1424-8220, www.mdpi.org/sensors

Shao Yang, Taff Gregory N, Walsh Stephen J (2011). Shadow detection and building-height estimation using IKONOS data. International Journal of Remote Sensing Volume: 32 Issue: 22 Pages: 6929-6944 DOI: 10.1080/01431161.2010.517226

Taff GN (2005) Conflict between local and global land-use values in Latvia's Gauja National Park. Landscape Research, 30(3):415-430

Taff GN, Müller D, Kuemmerle T, Ozdenerol E, Walsh SJ (2010) Reforestation in Central and Eastern Europe After the Breakdown of Socialism. In: Nagendra H, Southworth J, (eds.), Reforesting Landscapes: Linking Pattern and Process, Springer Netherlands (Landscape Series, volume 10, pp. 121 – 147

Turner, B. L., W. C. Clark, R. W. Kates, J. F. Richards, J. T. Mathews, and W. B. Meyer, eds. 1990. The Earth as Transformed by Human Action. Cambridge: Cambridge University Press.

Turner M, Wear D, Flamn R (1996) Land ownership and land-cover change in the Southern Appalachian Highlands and the Olympic Peninsula. Ecological applications, vol. 6. No. 4, 1150-1172

U.S. Geological Survey, Department of the Interior. 1996. Standards for digital orthophotos: National Mapping Program technical instructions

Vogelmann JE, Howard SM, Yang Y, Larson CR, Wylie BK, Driel JN (2001) Completion of the 1990's National Land Cover Data Set for the conterminous United States, Photogrammetric Engineering and Remote Sensing 67:650-662

Wood CH, Skole D (1998) Linking satellite, census, and survey data to study deforestation in the Brazilian Amazon., In: Liverman, DM (ed) People and pixels: Linking remote sensing and social science, National Academy Press, Washington, DC, pp. 94–120

World Bank Indicators (2010) Latvia Agricultural Production http://www.tradingeconomics.com/latvia/food-production-index-1999-2001--100-wb-data.html. Accessed 12 April 2012

Zvidrins P (1998) Changes in living standards and depopulation in Latvia in the 1990s. Social Indicators Research, vol. 43, No. 1 - 2, pp. 121-140

Chapter 3

Using Logistic regression to model the spread of invasive Giant Hogweed in Latvia

Introduction

Invasive species are a global problem with widespread effects on agriculture, fisheries, human health, and natural ecosystems (Andersen et al. 2004). In many areas, large scale colonization by non-native plants is changing nutrient cycling, increasing fire severity, and seriously compromising ecosystem condition and native biological diversity (Vitousek et al. 1996; Mack et al. 2000; Mooney and Cleland 2001). In this study "invasive species" refers to non-native species that become established in new locations, spread, and then cause ecological or economic harm or threaten human health (U.S. Fish and Wildlife Service 2003; Pimentel et al. 2000). The impacts of invasive species on plants and animals range from local suppression of single native species to species extinction and wholesale changes in the functioning of ecosystems (Ota 1993). Ecosystem functioning changes includes, among others: altering natural water resources, carbon sequestration, and biodiversity. Almost half of the US species listed under the Endangered Species Act are threatened by competition with invasive species (Wilcove et al. 1998) while, globally, invasive species are a major threat to 30% of birds, 11% of amphibians, and 8% of mammals on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (Baillie et al. 2004). Invasive species, once fully established may be very expensive and difficult to control and eradicate. For example the annual cost to the US economy to monitor, contain, and control invasive species is estimated to be between \$100 billion and \$200 billion (NASA 2009). In addition, homeowners spend an estimated \$500 million a year and golf courses spend \$1 billion a year to control non-native invasive species. Crop losses cost \$26 billion a year and an additional $3 billion is spent annually on herbicides to control non-native species (JFNew 2009). A fundamental approach to understanding and managing invasive species is to determine their current and potential distributions (Allen et al. 2006). Additionally, it is important to identify and minimize land uses that promote invasion, for example emplacement and improvement of roads (Forman and Alexander 1998; Trombulak and Frissell 2000). Roads provide dispersal of exotic species via three mechanisms: providing habitat by altering conditions, making invasion more likely by stressing or removing native species, and allowing easier movement by wild and human vectors.

Spatial modeling is a promising approach to predicting risk of invasion. Applying a predictive distribution model within a spatial context relies on the existence of landscape-scale variables that define suitable habitat for a species (Osborne, Alonso and Bryant 2001; Austin 2002a,b) based on the biological/ecological needs of that species. Spatial patterns of invasion can be predicted by linking current presence and absence of invasive species to spatially explicit predictor variables, like land use, geomorphology, and topography, using geographic information systems (GIS); (Store and Kangas 2001). Successful modeling efforts have demonstrated that establishing such spatial relationships requires extensive field data (Bethany et al. 2006). For example, Larson et al. (2001) collected data from more than 1300 transects in the Theodore Roosevelt National Park, North Dakota, USA to determine non-native plant relationships to native plant communities and anthropogenic disturbance. However, the paucity of relevant scientific data often makes it difficult to develop the type of predictive model that is required for decisions concerning invasive species management. Three types of data were used in this project: Giant Hogweed locational data collected from Public Participation Geographic Information Systems (PPGIS) involving Latvian high school geography students using GPS units; Giant Hogweed point data collected in the field by the author with the research team; Data of Giant Hogweed obtained from the Latvia ministry of Agriculture. These data (Giant Hogweed presence data) and pseudo absence data of Giant Hogweed generated using stratify random sampling were analyzed in a GIS and then input into a logistic regression for modeling and assessment. Logistic regression is a robust statistical tool to model species distribution (Smith 1994, Brito et al. 1999). Logistic regression models based on topography and other environmental factors have been used to predict species distribution, e.g. Charlotte et al. (2008) predict the distribution of invasive alien *Heracleum mantegazzianum* in Denmark using logistic regression, Zimmermann and Kienast (1999) predicted the distribution of alpine grasslands in Switzerland, and Robertson et al. 2003 used logistic regression to predict the potential distribution of species in South Africa.

In this study a logistic regression approach was used to describe the relationship between Giant Hogweed distribution and geographic and demographic features. The main objectives of this research are 1) to identify the most important variables influencing the distribution of Giant Hogweed in Latvia, 2) Validate the performance of the model and 3) produce a habitat suitability map of northeastern Latvia.

Background

Sosnowskyi Hogweed (*Heracleum sosnowskyi)* is a biennial or perennial herb in the carrot family (Apiaceae) which can grow to 12 feet or more (fig. 1a). Its hollow, ridged stems grow 2-4 inches in diameter. Its large compound leaves can grow up to 5 feet wide (fig. 1b). Its white flower heads can grow up to 2 1/2 feet in diameter. There are other poisonous Hogweed species related to Sosnowskyi Hogweed (*Heracleum mantegazzianum*, *Heracleum persicum*) that occur in many parts of the world, though all are native only to the Caucuses region. The *Heracleum sosnowskyi* species is a biennial or perennial plant mostly found out of its native range in the Baltic States, and especially Latvia. It is an invasive and highly toxic weed that threatens biodiversity, ecological health, and human health of infested areas in the Baltic countries (Estonia, Latvia, and Lithuania), northwest Russia, Belarus, Poland, and Ukraine even though the plant is most common in Latvia (Kabuce 2006). Local residents are afraid and have very special concerns about the safety of their children because the plant causes phytophotodermatitis (severe burns), painful blistering, permanent scarring and blindness when the sap of the plant comes in contact with the human body and is exposed to sunlight. The plant usually flourishes during summer and grows to heights of about 4-5 m (Pysek et al. 2007). Its large inflorescence is composed of many small flowers which can be white or sometimes pinkish. Flowering typically last from June to August in Latvia, and the seeds are egg-shaped or oval (Nielsen et al. 2005). Giant hogweed dies back during the winter months, leaving bare ground that can lead to an increase in soil erosion on riverbanks and steep slopes. The plant is hardy and can thrive in a cold climate (Kabuce, N. 2006). It was promoted as a crop for cattle feed in northwest Russia, where it was first introduced in 1947. From the 1940s onwards, it was introduced as a fodder plant to Latvia, Estonia, Lithuania, Belarus, Ukraine and the former German Democratic Republic (Nielsen et al. 2005). *H. sosnowskyi* was sown as a fodder plant for the first time in Latvia in 1948 and was grown on experimental agricultural farms (Gavrilova and Roze 2005). In addition to use as a crop for livestock feed, it was cultivated in many Botanic Gardens sometimes as an ornamental. In the 1960`s larger scale cultivation for forage needs began (Laiviņš and Gavrilova 2003). Continued plans to harvest hogweed commercially were halted shortly thereafter because its anise-like scent affected the flavor of the meat and milk of the animals that ate it and also posed health risk to humans and cattle (Nielsen C. et al. 2005). Giant hogweed quickly spread out from areas in which it was cultivated and began to populate the surrounding countryside. Currently, *H. sosnowskyi* in the Baltic States, especially in Latvia, has developed stands of hundreds and thousands of square meters, and the process of naturalization is intense (Gavrilova 2003). Giant hogweed grows on waste grounds, abandon fields (figure 1c), and wet areas along streams and rivers (fig. 1d), near houses, in vacant lots, and along railways and roads. It prefers moist soil and can quickly dominate ravines and stream banks. In Latvia *H. sosnowskyi* species is mostly found in artificial habitats (roadsides [figure 1e], disturbed areas [figure 1f], agricultural fields, abandoned farm yards and gardens) and semi natural habitats (bushes, grasslands, parks, pastures, abandoned orchards) (Gavrilova 2003). Giant Hogweed is an aggressive competitor. Because of its size and rapid growth, it out-competes native plant species by providing shade for native plants that are much in need of sunlight (Nielsen et al. 2005). Giant Hogweed overwhelms native species in occupied territories and therefore Giant Hogweed societies are poor in biodiversity (Gavrilova 2003). Presently, the naturalization of *H. sosnowskyi* is out of control and the plant has spread over almost all of the Baltic countries, mainly in unmanaged land areas and near ditches (Kabuce et al. 2010). Its naturalization is favored by abandoned land, particularly abandoned agriculture (Oboļeviča 2001), and also due to political changes (fall of communism) in the late 1980's and early 1990's, when Latvia experienced a shift in economic energy that refocused its workforce from an agriculture base into a new, service-oriented economy that forced many farmers to abandon their farms (Taff et al., 2010). According to Oboļeviča (2001), of the Latvia University of Agriculture, *H. sosnowskyi* has become so widespread in Latvia that it has outstripped the government's current capacity to eradicate or even control the invader.

The mode of dispersal of the plant is by both natural and human activities. The seeds are dispersed by running water, floods and by wind. Humans also carry the seeds on automobile wheels along roads, and in some cases the seeds can stick to clothes and be carried to other locations. The most common mode of dispersal is the establishment of new Giant Hogweed plants close to existing stands. A study by Pysek et al. (2007) showed that 90% of the seeds of the related *H. mantegazzianum* fall within a 4-meter radius of the plant.

Study Area

The study was carried out in Latvia which lies on the Baltic coast, in the western part of Eastern Europe. Latvia is one of the three Baltic States, situated on the east side of the Baltic Sea, the others being Estonia (to the north) and Lithuania (to the south). Latvia also borders Russia and Belarus to the east (fig. 2). The total land area is 64.6 thousand km^2 and the terrain is mostly low plain, with the majority of the territory between 40-200 meters above sea level (Eberhards 1984). The climate is wet with moderate winters for this latitude. The average amount of precipitation is 600-650 mm annually, the vegetation period usually lasts for 180-200 days (Normunds 1993). The landscape is characterized by matured forests, secondary forest, meadows, farmland, abandoned farmlands, lakes, rivers, hills, plains, villages and dispersed rural homesteads (Bunkse 2000).

Some land cover types are more vulnerable to invasion by Giant Hogweed than others e.g. cleared forest areas and abandoned agricultural areas are more susceptible to Giant Hogweed invasion than forested areas. Once an area is invaded by Giant Hogweed it develops large stands and dominates native species in occupied territories. It is easy to distinguish it from the surrounding vegetation especially during summer when the plant flowers. Large patches of Giant Hogweed are common to find in rural areas where agricultural abandonment is common due to out migration of workers into bigger cites in search for better opportunities. Latvia had the fastest growing economy in Europe from 2003 until the middle of 2008 (Eurostat 2010). After this period Latvia experience an economic setback which forced many Latvians to migrate to other European countries in search of better jobs and this caused a decrease in population. Most of Latvia's export is from wood and agricultural products including wildlife (Central Intelligence Agency 2009). Therefore, maintaining the health of the ecosystems in Latvia is important to policy and decision makers at all levels of government. The challenges caused by invasive species in general are numerous. They impact range site productivity, disturb wildlife habitats, and reduce biodiversity. During the field trip to Latvia in the summer of 2009, 2010 and 2011, the field team collected field data of Giant Hogweed locations and other land cover types using GPS in Valmiera, Cesis and Madoma located in the Vidzeme region in the country's northeast. This region was chosen because it is where the species was first introduced in the country as cattle fodder and records of the study species in this region are widely distributed

Fig. 1 Study area - northeastern Latvia

Methodology

Data collection

The data used were a LANDSAT Thematic Mapper (TM) image (path 186, row 20) obtained in the summers of 1992 and 2007 of northeast Latvia (downloaded from the USGS Global Visualization Viewer (Glovis)). The part of the image that corresponds to the study area covers approximately 6137 rows and 5515 columns. Digital elevation Model (DEM) and soil data were downloaded from DIVA-GIS (a free computer program for mapping and geographic data analysis). The LANDSAT TM image had a resolution of 30m x 30m. The DEM and soil data had a resolution of 90m x 90m and were resampled to 30m x30m resolution using the nearest neighbor technique to match the resolution of the LANDSAT image. GIS topographic layers of road network, urban centers, railroads, water, and protected areas were obtained from SIA Envirotech (Envirotech LTD). All data layers were converted to a 30m x 30m resolution for analysis. Giant Hogweed location points and pseudo-absence points were generated randomly. Ground-based surveys and monitoring by field personnel, including the author, with GPS equipment were used in summers of 2009, 2010 and 2011 for ground truth, particularly for satellite image classification and accuracy assessment purposes. This research team began ground control and ground truth in a preliminary field trip in the summer of 2009 led by Dr. Gregory Taff. In the summer of 2010 the research team collected data of Giant Hogweed in the form of public participation Geographic Information systems (PPGIS) involving Latvian geography high school students. PPGIS brings the academic practices of GIS and computer mapping to the public in order to promote knowledge production and potentially improve data quality and/or quantity. Public participation has the potential to accumulate large amounts of long-term data required to construct predictive models, which may otherwise be difficult to collect (Kadoya and Washitani 2007). In order to maximize probability of detection, surveys were timed to coincide with the growing season from June to August (Kabuce 2006). In addition the research team trained high school students to participate in the process of field data collection of Giant Hogweed locations in the form of public participation GIS. We also obtained locational data of Giant Hogweed presence (1969-2010, with a vast majority of cases in recent years since 2000) from the Latvian Ministry of Agriculture. These data together with PPGIS data, and data collected by the research team were used as Giant Hogweed presence points to run a regression model to predict suitable habitat of Giant Hogweed occurrence. As might be expected there are some limitations when data from various sources are pooled (Graham et al. 2004), e.g. collection bias in geographical space. We have more records near roads, towns and cities and fewer records in remote areas. For species distribution modeling, spatial bias in the record may not be a problem if the data are not environmentally biased (Newbold 2010). It is difficult to infer absence data; our data was presence only and pseudo-absence data need to be created for some modeling techniques (Graham et al. 2004) as in our case, requiring both presence and absence data. These issues were addressed by carefully checking the records and eliminating misleading data. Since true-absence data was not available, pseudo-absence data (Graham et al. 2004) was generated using methods to be discussed later in this paper. The Giant Hogweed location dataset was split into training data (70%) and test data (30%).

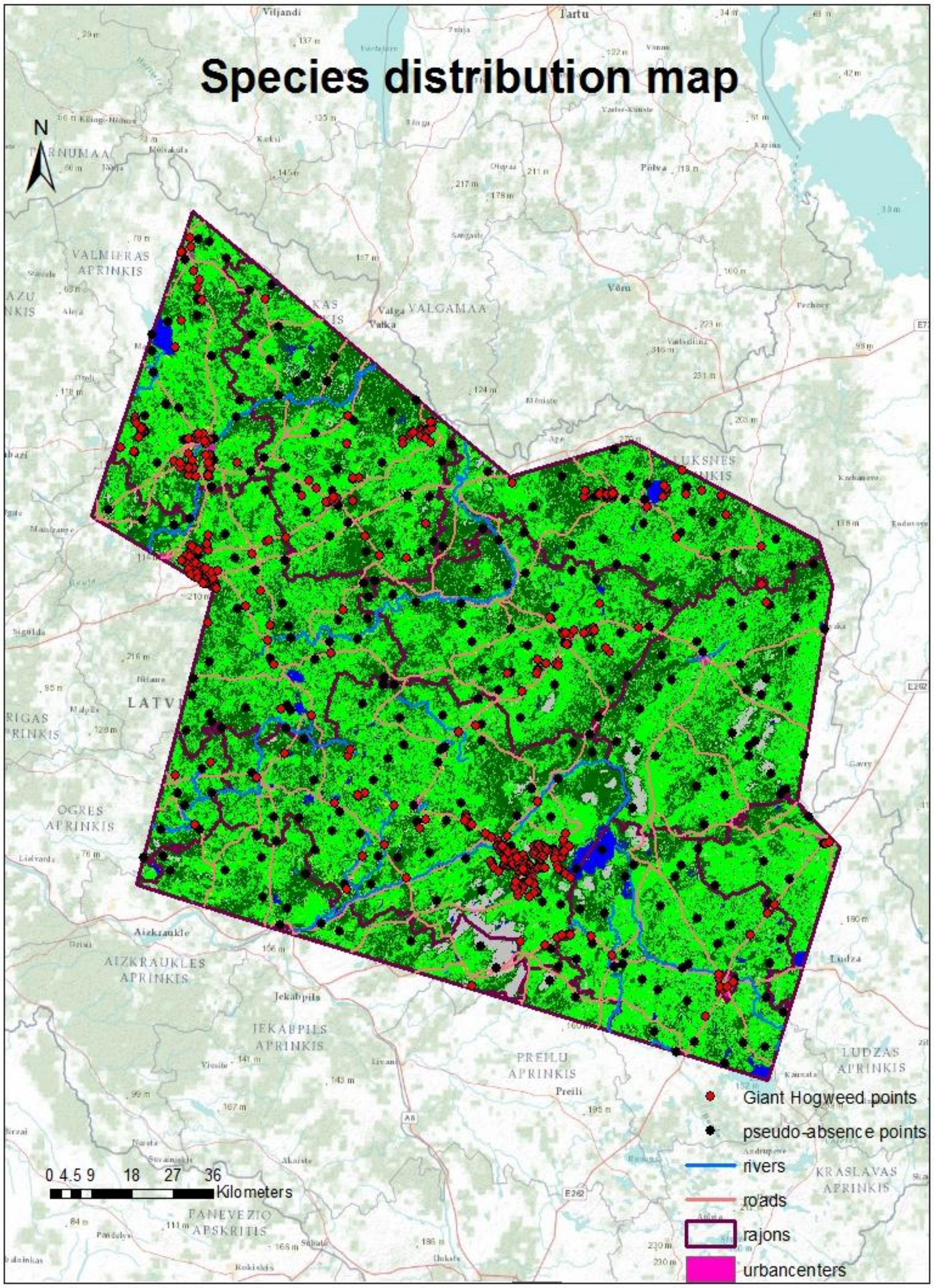

Fig. 2 Distribution of Giant Hogweed in northeastern Latvia using presence and pseudo-absence data

Environmental Variables

The distribution of species in space and time is highly influenced by biological and anthropogenic factors including their environmental context (Liu et al. 2011). Environmental variables are potentially related to species distribution and may have a direct or indirect impact on their distribution. Understanding the main drivers shaping species distribution is particularly important for species distribution modeling (Pulliam 2000). Even though many variables may be related to species distribution it is important to include only those that are ecologically relevant to the target species (Elith and Leathwick 2009).

In this study ten explanatory variables (Table 1) were generated in ESRI ArcGIS 10, to represent resource availability and topographical features on the landscape that may facilitate invasion (Davis et al. 2000). Most of these variables were chosen based on extensive literature review. Many studies have focused on identifying factors that influence invasibility. For example Davis et al. (2000) showed that topographic features (elevation, slope, and aspect) on the landscape may facilitate invasion. Some researchers have evaluated the relationship between disturbance regimes and invasibility. Greenburg et al. (1997) and Tyser and Worley (1992) found that roadways and trails facilitated non-native species spread. Other studies have found that riparian areas were more invaded than upland sites (Stohlgren et al. 1998; Levine 2000; Larson et al. 2001). Soil type has also been proposed as a factor related to invasibility. Harrison (1999) determined that invasion was greater on soils with greater levels of nutrients. Human population density was correlated with the distribution of *H. mantegazzianum* in the Czech Republic (Pysek et al. 1998).

Slope and aspect were derived from a Digital Elevation Model in the study site using Spatial Analyst in ArcGIS 10. These variables were derived at the original resolution of the DEM (90m) and then resample to 30 m using nearest neighbor technique to match the resolution of the LANDSAT image. The nearest neighbor technique was used because it tranfers the original pixel values which is essential for post-classification change detection. One limitation of this method is that the edges of the image may appear jagged. Euclidean distances from features (e.g. roads, railways, water bodies, and urban centers) were created from GIS shape files using the map calculator function in the Spatial Analyst extension in ArcMap. Soil and LULCC data were included in the analysis. All the data were converted to the same scale (30m x 30m).

Elevation, aspect and slope were selected because they have been demonstrated to be useful surrogates for the spatial and temporal distribution of factors such as radiation, precipitation, and temperature that influence species composition and productivity (Albert et al. 2007). Roads, railways, and land-cover change were selected because human disturbance is very important for the expansion and persistence of many invasive species. Previous studies have demonstrated that these landscape characteristics often play an important role in species distribution and patterns of invasion (lambrinos 2002; Stohlgren et al. 2003; Kumar et al. 2006).

Table 1 Predictor variables

Variable	Description	Ecological meaning
Elevation	Altitude from sea level	Influence climatic conditions
Slope	Slope angle (degrees)	Stability, erosion, moisture
Aspect	North,Northeast, East, Southeast, South, Southwest, etc.	Solar radiation, wind
Soil	Unconsolidated mineral or organic material.	Nutrient availability, nutrient retention capacity, rooting conditions
Land cover change	Change to forest, agriculture, and urban	Affects the distribution of plants and animals
Distance to nearest road	Highways, paved roads	Influence plants and animal distributions
Water bodies	Rivers, streams, lakes	Influence plants and animal distribution
Rail roads	Rail roads	Influence species distribution
Distance to urban center	Central districts, high population	Humans influence species distribution

	areas	
Presence of protected areas	Presence, absence	Influence plants and animal distributions

Modeling Technique

Model selection

Several ecological niche models have been widely used to predict potential geographic distribution of non-native species, such as bioclimatic prediction system (BIOCLIM) (Busby 1991), generic algorithm for rule-set prediction modeling system (GARP) (Stockwell and Peters 1991; Stockwell et al. 2006), artificial neutral network (ANN) (Pearson et al. 2002), ecological niche factor analysis (ENFA) (Hirzel et al. 2002) and regression methods such as generalized linear model (GLM) (Lehman et al. 2002), generalized additive model (GAM) (Elith et al. 2006), boosted regression trees (BRT) (Leathwick et al. 2006) and multivariate adaptive regression splines (MARS) (Elith et al. 2007). The logistic regression model was selected for this study and was preferred to other regression methods because it is 'terse', the dependent variable (Giant Hogweed presence/absence) is a dichotomous variable, and logistic regression has already proved its value in many case studies (Robertson et al. 2003; Brenning 2005).

Logistic regression model is a type of Generalized Linear Model (GLM) appropriate for binary outcome data such as species presence or absence (McCullagh and Nelder 1989; Evans et al. 2000). For example, David et al. (2009) used logistic regression to model the potential spread of invasive plants in the Green Ridge State Forest in Western Maryland along the Potomac River. Site features that have been associated with invasibility include both environmental and anthropogenic factors such as disturbance (Almasi 2000; Silveri et al. 2001), proximity to roads (Harrison et al. 2002), soil nutrients, topographic position, and forest fragmentation (Brothers and Spingarn 1992; Cadanasso and Pickett 2001). A logistic regression model was built and run in SPSS IBM statistic 20 (IBM 2011) using distance to road edge, distance to river edge, distance to urban centers, soil, LULCC, elevation, aspect, slope and protected areas as predictor variables, and Giant Hogweed presence/absence as the outcome variable to determine which variables were significant, and to assess the potential habitat distribution of Giant Hogweed in the study area.

A logistic regression model of the type below was used:

$$\mathrm{logit}(p) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \ldots + b_k X_k$$

Where, p is the probability of presence of the characteristic of interest (presence of Giant Hogweed), b_o is a constant, $b_1,\ldots,b_k$ are the model coefficients $X_1,\ldots,X_k$ are the independent variables.

$$\mathrm{logit}(p) = \ln\left[\frac{p}{1-p}\right]$$

and

$$\mathrm{odds} = \frac{p}{1-p} = \frac{\textit{probability of presence of characteristic}}{\textit{probability of absence of characteristic}}$$

The outcome variable will be 1 if Giant Hogweed exists at that location and 0 for locations having no Giant Hogweed.

The logistic regression model is widely used with presence and absent data but has some disadvantages in that absence data is needed (absence of species does not necessarily mean the habitat is not suitable for the species but could be due to natural barriers that have prevented the species from habiting that area and logistic regression is also sensitive to multicollinearity (McCullagh and Nelder 1989). Logistic regression requires both presence and absence data. Data on Giant Hogweed locations obtained from the research team, PPGIS, and the Latvian Ministry of Agriculture were used as presence data, and absences were generated as pseudo-absences generated at random with a geographic weighted exclusion (points within a given buffer distance will be excluded) (e.g. Hirzel et al. 2001). Based on the author's field work and the size of Giant Hogweed patches in the field together with literature review, a distance of 500m was chosen to reduce the probability of selecting pseudo-absence points that are actually presence points. A 500m buffer around each present point was created; the absence points were then randomly drawn from the areas out of these buffers, a procedure similar to that used by Akcakaya Atwood (1997). In total there were 476 presence points and 364 pseudo-absence points.

The logistic regression was run using the stepwise procedure, i.e. enter significant variables sequentially; after entering a variable in the model, check for multicollinearity among the explanatory variables and possibly remove variables that became non-significant (p-values > 0.10). Significance of the predictors in the logistic regression model was assessed at the 0.05 and 0.01 levels. The data was divided into training (70%) and test (30%) data using random selection and the model was run using stepwise logistic regression on the training data. The overall accuracy (proportion of Giant Hogweed points that were correctly classified as 1 (present) and 0 (absence)) for the training and test data were very similar 71% (training data) and 70% (test data), showing the data fit the model well (table 2). Due to the similar results between testing and training data, it was deemed beneficial to run the model again with all the data to increase the sample size and test the null hypothesis to see if the independent variables have any effect on the dependent variable, as shown in table 3 & 4.

Model Validation

The model evaluation is an important step in modeling and, ideally, models should be tested with independent data (Guisan and Zimmermann 2000). There are several methods of assessing model performance. The first approach uses test data that are collected independently from the data used to calibrate the model (Fleishman et al. 2002). The second approach partitions the available data into calibration and test datasets (e.g. Pearson et al. 2002). As stated, the second approach was used in this study because of distance and cost from the U.S. to Latvia to collect independent test data. The datasets were divided into training (70%) to develop the model and the other 30% (test data) was used to evaluate the model.

The predictive successes of the model's performance were evaluated by the overall accuracy from the confusion matrix for both the training and the testing data. The receiver operating characteristic (ROC) technique was also used to evaluate the performance of the model. The use of this threshold-independent method has increased in ecological applications in recent years (Osborne, Alonso and Bryant 2001). The ROC curve is a plot of true positive cases (a true positive in this study represents a predicted value of greater than 0.5 for Giant Hogweed presence where Giant Hogweed was actually present) against corresponding false positive cases (a predicted value of greater than 0.5 where Giant Hogweed was absent). The calculation of area under the curve (AUC) of the

ROC plot provides a measure of discrimination ability that varies from 0.5 – 1. The bigger the value of AUC is, the stronger the relationship between the presence locations and the environmental variables, which indicate better performance of the model. For the logistic regression, the software SPSS was used for ROC analysis and calculations of the AUC is based on cut-off values taking into account the full range of threshold values. AUC values > 0.7 indicate a good fit of the model to the original data.

Results

Table 2 Stepwise logistic regression using training (70%) and test data (30%)

Observed			Predicted					
			Selected Cases[b]			Unselected Cases[c]		
			presence/absence		Percentage Correct	presence/absence		Percentage Correct
			0	1		0	1	
Step 1	presence/absence	0	167	90	65.0	71	36	66.4
		1	75	260	77.6	34	107	75.9
	Overall Percentage				72.1			71.8

a. The cut value is .500

As mentioned, since the outcomes were so similar between the training and testing data (a test to show that that data fits the model), it was deemed useful to re-run the analysis using all the data as training data in order to increase the sample size.

Table 3 The model assumes independent variables do not have any effect on the dependent variable (null hypothesis).

Block 0: Beginning Block

Classification Table[a,b]

Observed	Predicted

			presence/absence		Percentage Correct
			0	1	
Step 0	presence/absence	0	0	364	0.0
		1	0	476	100.0
	Overall Percentage				56.7

Table 4 Classification Table using all the data set. Model assumes independent variables have an effect on the dependent variable.

Observed			Predicted		
			presence/absence		Percentage Correct
			0	1	
Step 1	presence/absence	0	224	140	61.5
		1	96	380	79.8
	Overall Percentage				71.9

Table 5a Results from the ENTER logistic regression including all the variables that contributed to model performance with the exception of slope.

Variables in the Equation

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1[a]	Distance to road (km)	-.631	.132	23.003	1	.000	.532
	Distance to road (0.1km)	1.080	.249	18.835	1	.000	2.945
	Distance to railway (0.3km)	1.538	.577	7.090	1	.008	4.653
	Distance urbancenter (km)	-.036	.013	7.501	1	.006	.965
	Distance to water (0.3km)	.492	.183	7.252	1	.007	1.636

slope	.073	.051	2.061	1	.151	1.076
LULCC (change to forest)_3x3			14.174	3	.003	
LULCC (change to forest)_3x3(1)	-.445	.665	.448	1	.503	.641
LULCC (change to agriculture)_3x3(2)	-1.542	.724	4.542	1	.033	.214
LULCC (change to urban)_3x3(3)	-.915	.743	1.516	1	.218	.401
soil			11.909	2	.003	
soil(1)	.517	.283	3.343	1	.067	1.676
soil(2)	-.143	.314	.208	1	.648	.867
Elevation	-.009	.002	19.668	1	.000	.991
Distance to water (km)	-.071	.034	4.413	1	.036	.932
Distance to urban center (5km)	.026	.283	.008	1	.927	1.026
Constant	2.289	.785	8.511	1	.004	9.866

Table 5b Variables not in the equation includes all variables that did not contribute significantly to model performance after the last step (13) from the stepwise logistic regression.

			score	df	Sig.
Step 13	Variables	Aspect	.027	1	.870
		Distance to railway (0.015km)	.017	1	.896
		Distance to railway (0.02km)	.017	1	.896
		Distance to railway (0.05km)	.092	1	.762
		Distance to railway (0.1km)	1.508	1	.219
		Distance to railway (0.5km)	1.266	1	.261
		Distance to road (0.015km)	.057	1	.812
		Distance to road (0.02km)	.029	1	.866
		Distance road (0.05km)	.131	1	.717
		Distance to road (0.5km)	1.162	1	.281

Distance urbancenter (0.5km)	.436	1	.509
slope	3.464	1	.063
Distance to water (0.015km)	.355	1	.551
Distance to water (0.02km)	2.974	1	.085
Distance to water (0.03km)	1.605	1	.205
Distance to water (0.05km)	.137	1	.711
Distance to water (0.1km)	.256	1	.613
Distance to water (0.5km)	1.004	1	.316
Presence/bsence of protected areas	1.776	1	.183

The variables not included in the model are aspect, slope and percent protected areas (table 5b). These variables were insignificant. For some reason SPSS does not differentiate between variables and coded variables e.g. types of soil. If the variable is significant, then all the coded variables will be included irrespective of whether the individual coded variables are significant or not (soil is significant and soil(1), soil(2) and soil(3) are included in the model even though soil(1) and soil(2) are insignificant. Slope was added to the ENTER logistic regression model (table 5a) because it had value very close to significant and secondly slope has been shown to be an important factor affecting Giant Hogweed spread in other studies (Pysek et al. 2005).

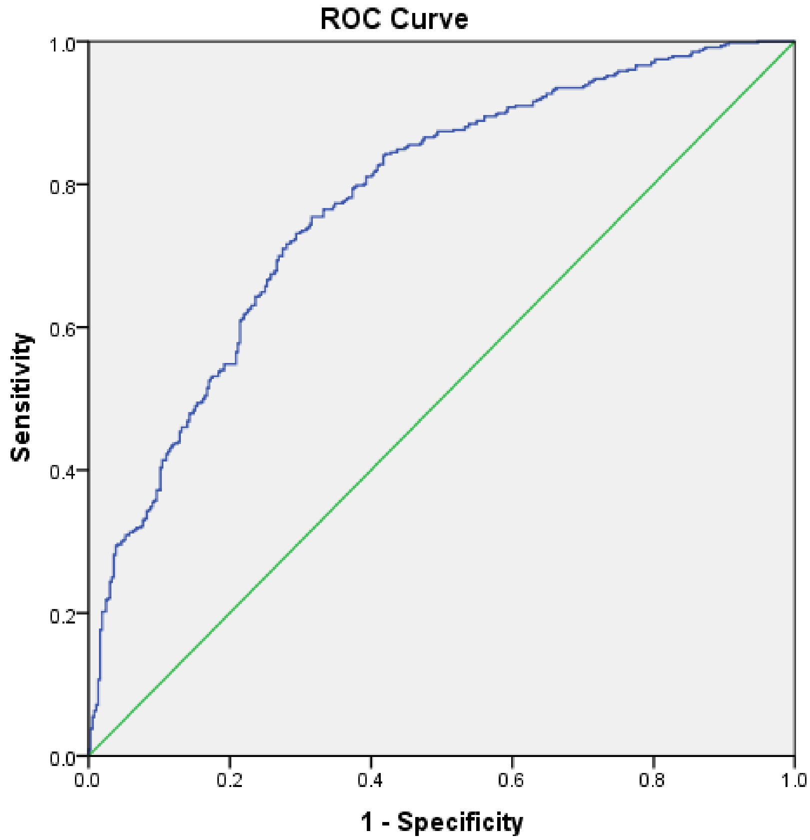

Fig. 3 Receiver Operating Characteristic Curve for the model performance. The Area Under Curve (AUC) is equal to 0.776 and the standard error for the AUC equals 0.016 (table 5). A larger AUC indicates better agreement with the data. AUC scale: 0.7-0.8 acceptable, 0.8-0.9 excellent, 0.9-1 outstanding (Hosmer and Lemesshow 2000).

Table 6 Area Under the curve

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval Lower Bound	Asymptotic 95% Confidence Interval Upper Bound
.776	.016	.000	.745	.808

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

From table 2, the overall percentage correctly classified for the training data was 72.1 % and for the test data was 71.8 %. These two values are close (the difference between the two values is less than 2 %) meaning the data are not overfitted to the model.

Accuracy measures for the logistic regression are reported using the classification table. The classification table is a table tells us how many of the cases where the observed values of the dependent variable were 1 or 0 respectively have been correctly predicted. In the Classification table (table 3), the columns are the two predicted values of the dependent, while the rows are the two observed (actual) values of the dependent. In a perfect model, all cases will be on the diagonal and the overall percent correct will be 100%. Accuracy measures are reported as overall percentage (number of cases correctly classified), specificity (percent of absences correctly classified); sensitivity (percent of true presences) and AUC (area under the curve).

Table 3 shows the overall accuracy for the model performance using all the data as training (71.9 %). This is a considerable improvement on the 56.7% correct classification with the constant model so we know that the model with predictors is a significantly better model than the only constant model. The percentage of Giant Hogweed that was correctly predicted as present by the model is 79.8 % and Giant Hogweed that was correctly predicted as absent is 61.5 %.

Importance of variables in the model

Stepwise logistic regression removes all variables that do not contribute to the model's performance and only retains those variables that contribute significantly to the model's performance. All utilized variables were entered into the stepwise logistic regression, though some were deleted from the model due to a lack of significance. Aspect and protected areas did not contribute to the model's performance and were therefore not included in the model. Seven of the eight topographic variables contributed significantly to the logistic regression model. Distances to roads, water, railways, and urban center were the most important factors determining the probability of occurrence of Giant Hogweed. Slope, LULCC and soil contributed to a lesser extent. For roads, rail roads and water, buffer distances of 0.015 km, 0.02 km, 0.03 km, 0.1 km, 0.3 km, 0.5 km were tested. For urban centers, buffer distances of 0.5 km, 1 km, 3 km, 5 km and 10 km were also tested. The probability of finding Giant Hogweed was predicted to increase at sites within 0.1km from roads and within 0.3km from water and railways. It is also more likely to find Giant Hogweed within 5km from urban centers and less likely further from than 5km. It is also less likely to find Giant Hogweed at higher elevations.

Only the contribution of slope was slightly insignificant (0.151). The included variables contributed significantly to the model according to the Wald statistics.

Habitat suitability map

Habitat modeling generated using spatial statistics and GIS can help in the characterizations of habitat requirements and the localization of suitable habitats (Guisan and Zimmermann 2000). Habitat distribution models or predictive distribution models are probability maps that depict the likelihood of occurrence of a species (Store and Kangas 2001). In this analysis, the logistic regression model output was used to predict probability of potential future Giant Hogweed invasion. In order to map the analysis results, the variable values at *each* pixel within the study site were calculated and multiplied by their respective parameter estimates from the model (Table 4) using the spatial analyst raster calculator in ArcMap 10. This procedure allocates a value for the logarithm of the odds (i.e. logit(P)) of detecting Giant Hogweed in all pixels in the study site (Fig. 4a) using the equation: Logit(P) = B0 +B1X1…….BkXk. Figure 4b shows a small zoom in area of possible Giant Hogweed colonization in relation to roads, railways, rivers and urban centers.

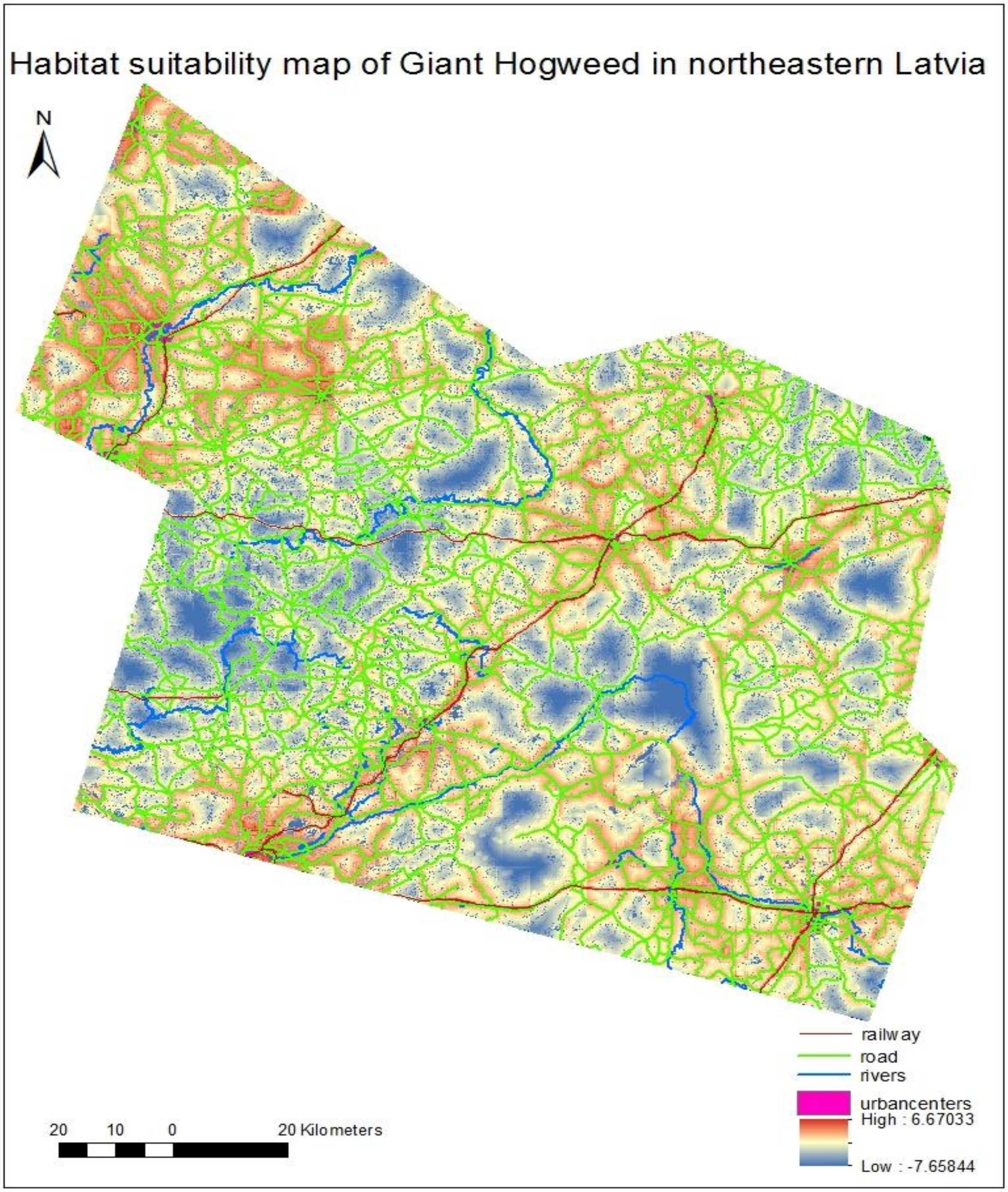

Fig. 4a Habitat suitability map of Giant Hogweed using the formula Logit(P) = B0 +B1X1…….BkXk.

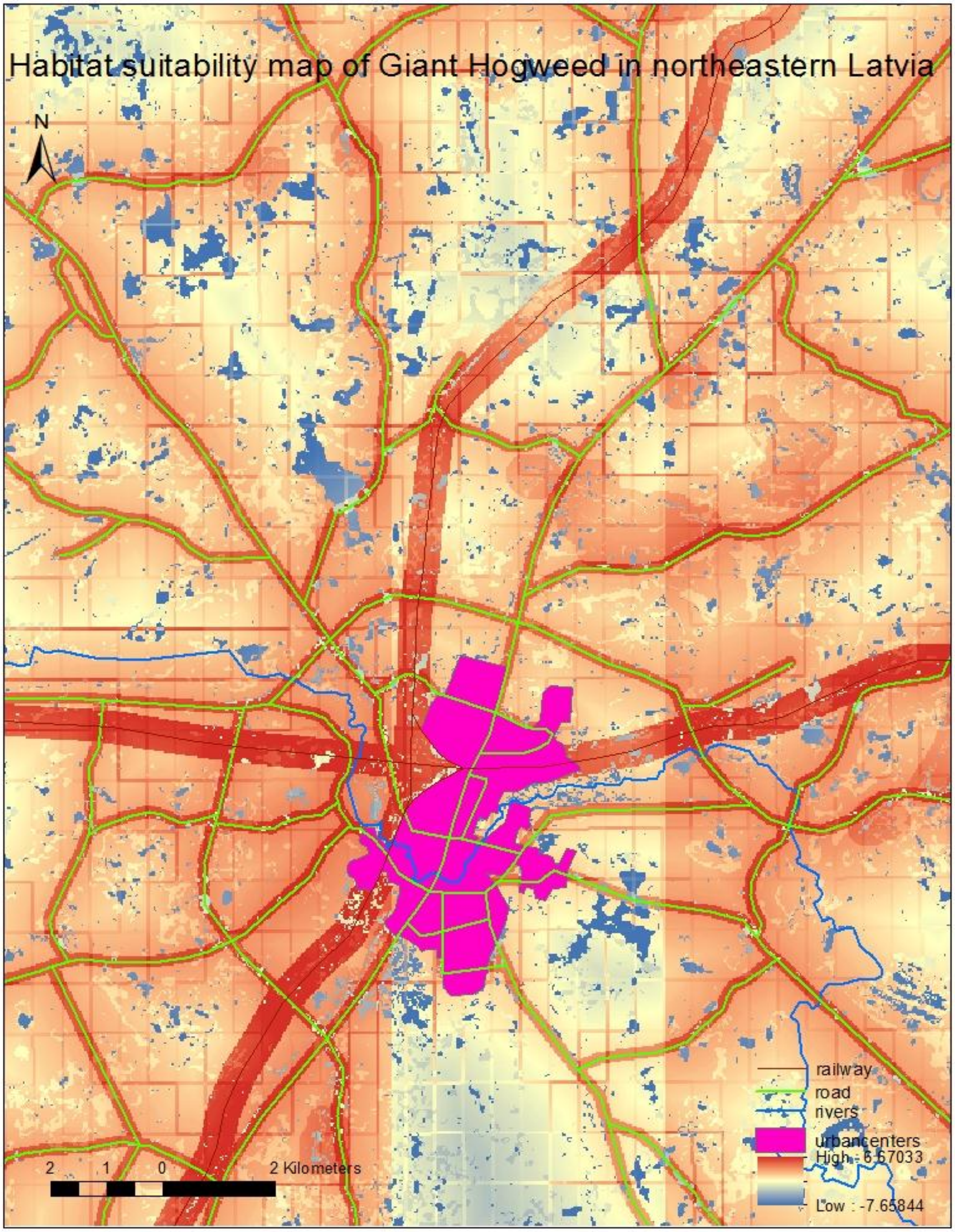

Fig. 4b A small zoom in area (habitat suitability map) showing possible Giant Hogweed colonization in relation to roads, railways, rivers and urban centers.

The logit value was subsequently converted into a probability using the formula P = exp(logit value)/(1+exp(logit value)) in the spatial analyst raster calculator. The resulting map shows probability values as suitability predictions (Fig. 5a). Fig. 5b is a zoom in area showing suitable areas of Giant Hogweed colonization in relation to selected variables such as roads, rivers and urban centers. Note that these are probabilities of Giant Hogweed suitability, though not direct probabilities of invasion.

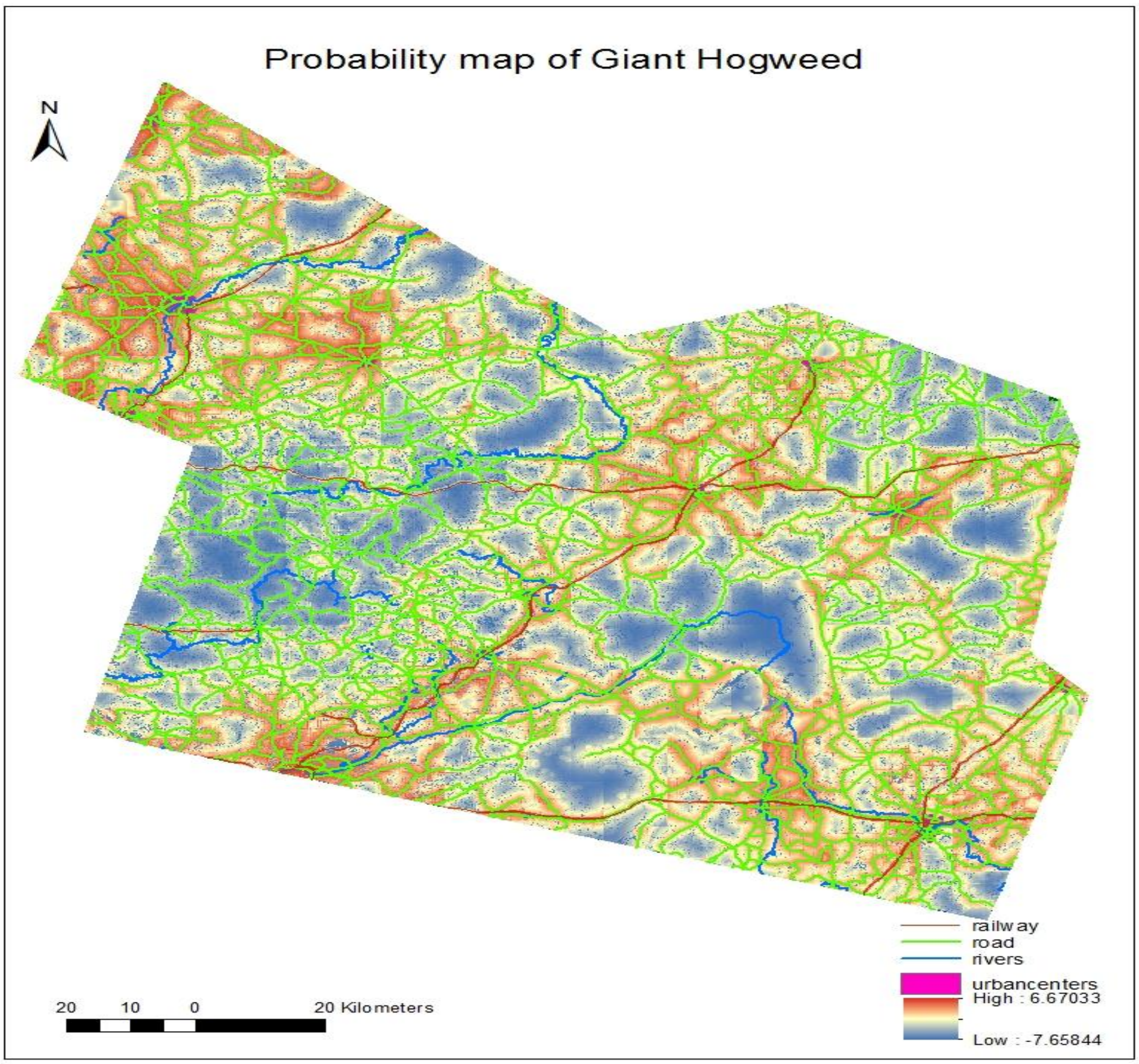

Fig. 5a Probability map of Giant Hogweed using the formula P = exp(logit value)/(1+exp(logit value))

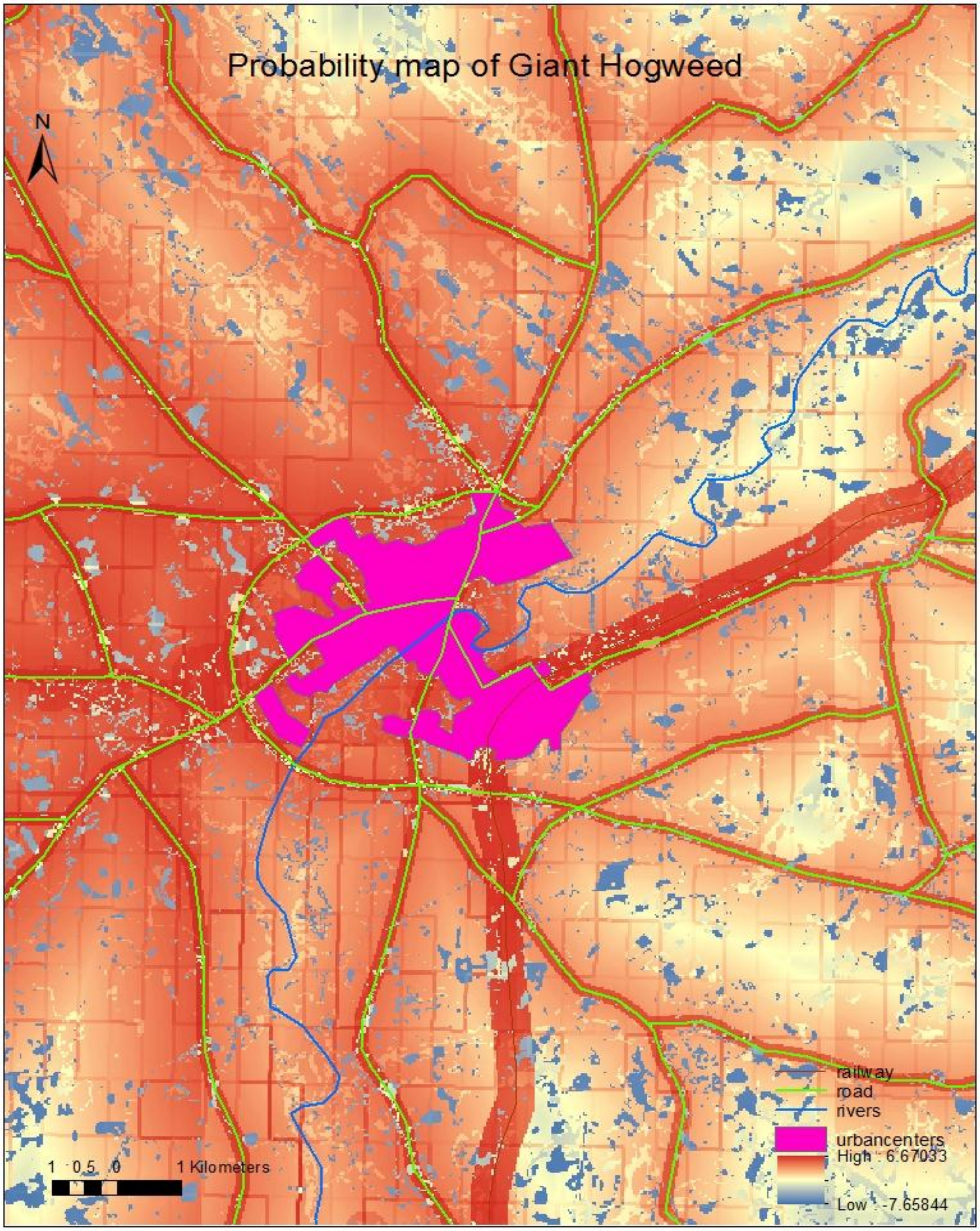

Fig. 5b A small zoom in area (probability map of Giant Hogweed) showing possible Giant Hogweed colonization in relation to roads, rivers and urban centers.

Threshold selection

Binary predictions of ‘presence’ or ‘absence’ are necessary to test model performance using statistics derived from the confusion matrix. That is, when you run the regression model there is an option to save statistics from the confusion matrix such sensitivity and specificity for all cases. The sensitivity and specificity values can later be used to generate a ROC curve from which a threshold value can be determined. The continuous suitability map therefore needed to be transformed into a binary output using a threshold value. A number of different methods have been used to select threshold occurrences which include fixed value (Manel et al. 1999; Robertson et al. 2001). This method is said to be subjective and lacks ecological reasoning (Liu et al. 2005). Another method is to use the lowest predicted value (Pearson et al. 2006; Phillips et al. 2006). This method assumes that species presence is restricted to locations equally. The Fixed sensitivity method (Pearson et al. 2004) allows for certain omission (5%) and is less sensitive to outliers. Other methods to select threshold values include: Sensitivity-specificity equality (Pearson et al. 2004), sensitivity-specificity sum maximization (Manel et al. 2001), Maximize kappa (Huntley et al. 1995; Elith et al. 2006), Average probability/suitability (Cramer 2003), and equal prevalence (Cramer 2003). In this research the sensitivity-specificity sum maximization was used as the threshold method for the following reasons:

1) The data included presence and absence data
2) The research is focused on predicting where Giant Hogweed occurs
3) If the threshold value is very low then sensitivity will be high and specificity will be low (can lead to over prediction) and if the threshold value is very high then specificity will be high and sensitivity will be low (can lead to under prediction).

4) Maximizing the sum of sensitivity and specificity assumes the two quantities are equally important and thereby reduces the chance of over prediction or under prediction of the model's performance.

5) Liu et al. (2005) tested twelve methods for setting thresholds using presence and absence data for two European plant species. Based on four assessments of predictive performance (sensitivity, specificity, accuracy and Kappa), they concluded that the best methods for setting thresholds included maximizing the sum of sensitivity and specificity.

Based on these factors we expect sensitivity to be a non-increasing function of the cut-off and specificity to be non-decreasing function of the cut-off. Fig 6 shows the value of the cut-off that simultaneously maximizes both the sensitivity and the specificity (the graph was constructed using saved probabilities from the ROC curve). In the diagram, this occurs where the two curves cross. At this point, the cut-off is estimated to be 0.567 and the sensitivity and specificity are equal to 0.72. In order to produce a binary prediction map of Giant Hogweed, values above the cut-off (0.567) were coded as 1 indicating Giant Hogweed presence and values below the cut-off were coded as 0 to indicate Giant Hogweed absence (Fig. 7).

To evaluate the binary map, known Giant Hogweed points were overlaid on the map as shown in Fig. 8. The percentage of Giant Hogweed points (from the testing dataset) that fell in areas that are considered suitable is 75%.

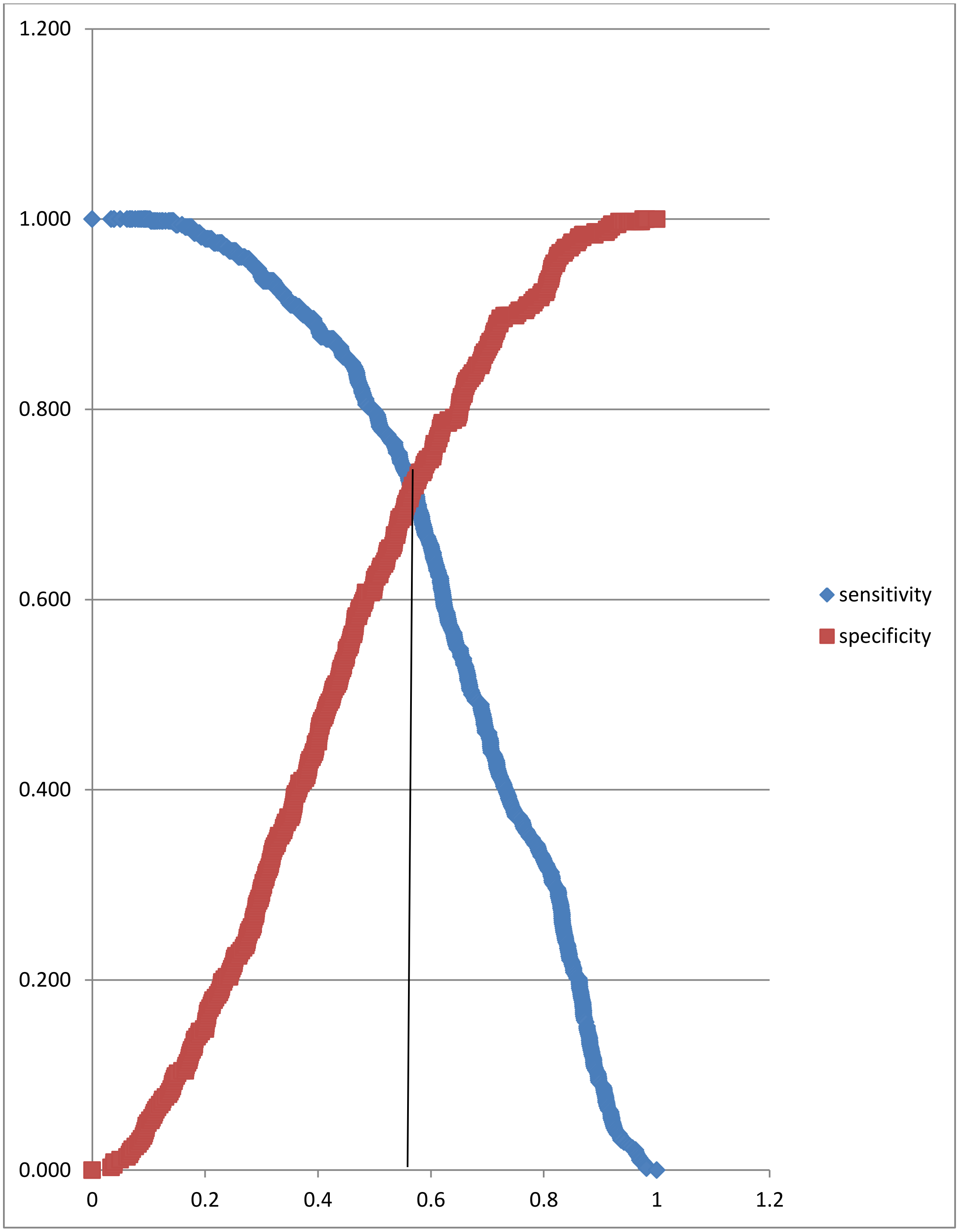

Fig. 6 Maximizing specificity and sensitivity

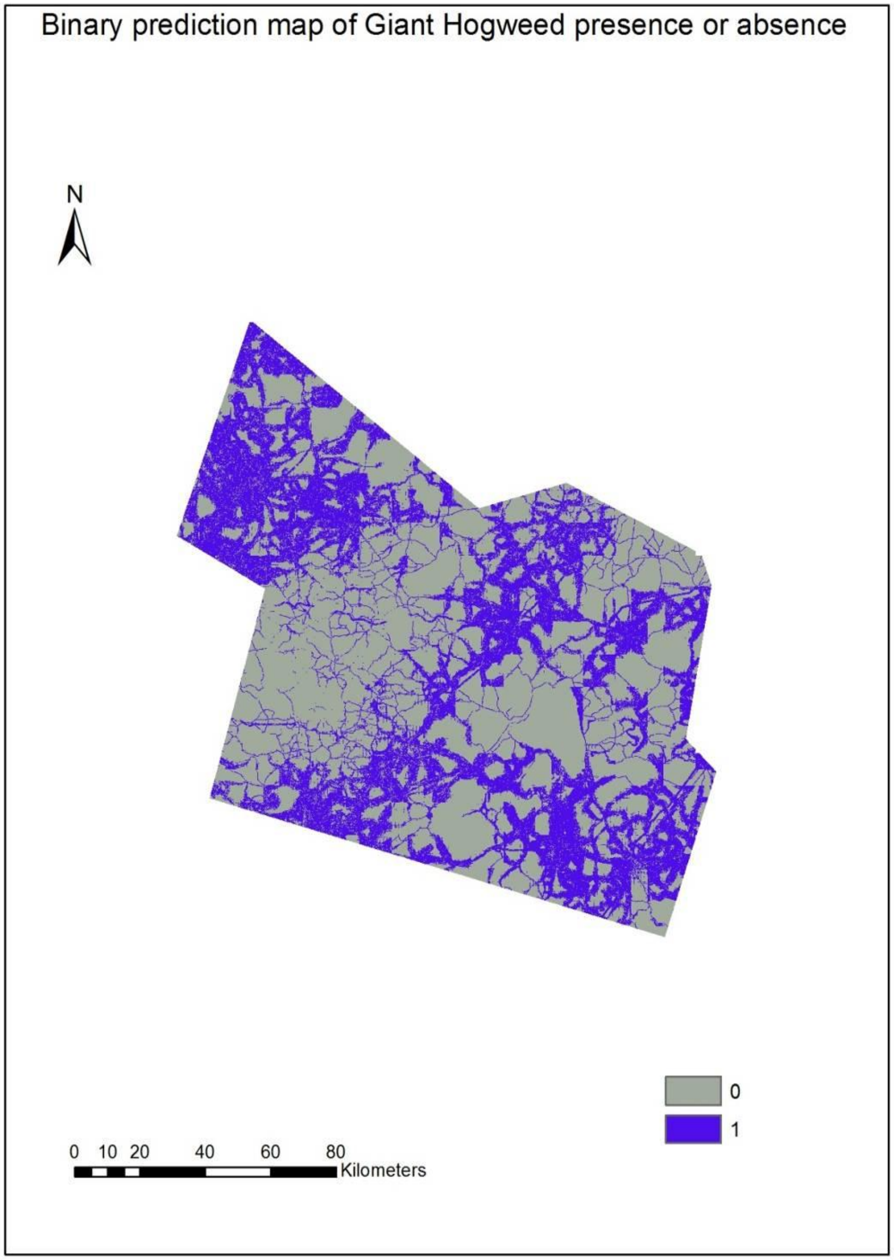

Fig. 7 Binary prediction map of Giant Hogweed (1= Giant Hogweed presence and 0 = Giant Hogweed absence)

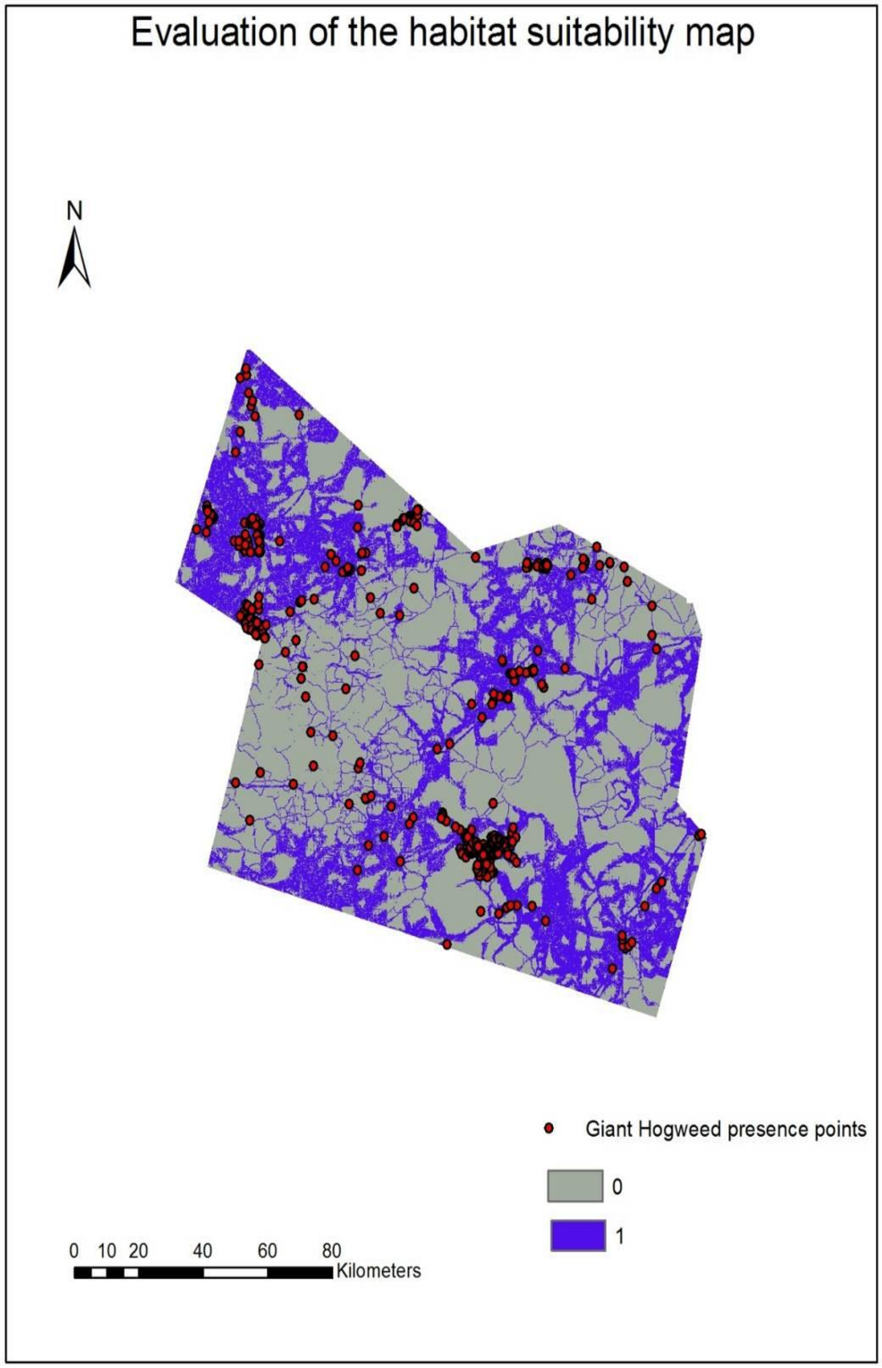

Fig. 8 Test of habitat suitability

Discussion

The results reported in the previous section (figure 4) show that distance to road, distance to water bodies and distance to urban centers are useful in predicting data on weed presence and absence. The negative correlations between these variables and Giant Hogweed presence indicates that the further away from roads, water bodies and urban centers, the less likely Giant Hogweed is to be found. This is verified by previous research that Giant Hogweed flourishes near roads, rivers and close to urban centers (Pysek et al. 2005). Giant Hogweed seeds are transported by running water, tires of vehicles, and by humans. Close to urban centers human activities increase and Giant Hogweed seeds are more likely to be transported. Giant Hogweed thrives well in abandoned fields and in Latvia many abandoned fields are closer to roads which increase the chances of finding Giant Hogweed near roads. Also further away from water bodies the soil gets dryer and Giant Hogweed thrives well in moist soil. The positive correlation between Giant hogweed presence and distance to railway is misleading. The reason is likely because there are very few railways in the study area (just five), and therefore many of the distances of pixels in the study site to railways are quite high. Topographic variables such elevation, slope, and aspect are also important in predicting the locations of Giant Hogweed. The negative correlation between elevation and Giant Hogweed occurrence indicates it is less likely to find Giant Hogweed at higher elevation. Even though slope is slightly insignificant, it was included in the model because other studies indicate that slope plays an important role in Giant Hogweed spread (Pysek et al. 2005). The results from the logistic regression show that Giant Hogweed is more likely to grow on higher slopes. The reason could be that on higher slopes there is more abandoned land, and therefore the land is more susceptible to Giant Hogweed invasion. Using the stepwise logistic regression, aspect was left out (did not contribute to the model prediction) possibly because Latvia is a relatively flat country (slopes quite low in general) and exposure to the sun is quite evenly distributed. Human disturbances such as land use land cover change also play an important role in the spread of Giant Hogweed. Land-cover change was coded as dummy variables with 1=change to forest, 2=change to agriculture and 3=change to urban. All three dummy variables were negatively correlated with Giant Hogweed spread. When land changes to forest, the forest canopy shades Giant Hogweed from receiving sunlight and therefore reduces its chances of surviving. When land is continuously used for agriculture or urban development, it prevents Giant Hogweed from spreading. During field trips to Latvia, anecdotal stories from farmers revealed that Giant Hogweed spread increases when a farm is abandoned for one or more years. Soil too was negatively correlated with Giant Hogweed occurrence.

Conclusion

The invasion dynamics of exotic species depend to a large extent on characteristics of the landscape (Planty-Tabacchi et al. 1996; Garcia-Robledo and Murcia 2005; Thomas et al. 2006). Landscape factors interact with biotic and abiotic factors such as human disturbance and climate. Together, these factors can allow exotic species to become abundant and persistent differentially within the landscape, and contribute to their potential nuisance or pest status (Rand et al. 2004; Knight and Reich 2005). Disturbance frequently favors invasion (Hobbs 1989; Hobbs and Huenneke 1992; Orians 1986). The distribution of Giant Hogweed in northeastern Latvia is strongly influenced by distance to roads, distance to water bodies, distance to urban centers, elevation, land-cover land-use change and, to a lesser extent, slope and soil type. Results from the logistic regression suggest that Giant Hogweed thrives well near roads, near water bodies, and proximity to urban centers in low elevation areas and moderate slopes. During the years of economic growth, there was a tendency to build family houses in the suburbs and commute to work in cities. The urban sprawl caused a reduction in natural areas and increased landscape fragmentation, and these problems still remain, even though the period of economic expansion is now over (European Environment Agency 2010). During this period transportation development also saw a rise to connect many Latvian cities to accommodate economic growth. This research strongly suggests that urban sprawl and road networks are important factors that influence the spread of Giant Hogweed. Maybe the most significant contribution of this study from a methodological stand point is that, for the first time it has been determined quantitatively from what distances from roads, rivers, rail roads and urban centers we are more likely to find Giant hogweed as opposed to just mentioning that Giant Hogweed thrives near roads, rivers, and rail roads in other studies.

This study also shows how land cover and LULCC affect the distribution of invasive Giant Hogweed in the study area. In recent years, Latvia has experienced a significant trend towards the abandonment of agricultural land, caused by the unfavourable economic situation in the agricultural sector, rural-to-urban migration and the ageing of the rural population (European Environment Agency 2010). The depopulation of rural areas leads to agricultural land abandonment. Open disturbed areas such as cleared forest or abandon agriculture are favorable areas for Giant hogweed to thrive. Therefore, one way to reduce Giant Hogweed spread presently and in the future is to plant trees where land is cleared and in abandoned farmlands. Giant Hogweed does poorly in closed canopy (forested areas) where there is less sunlight and in lands that have continuously been used for agriculture or development.

Plant traits favoring invasion include lack of controlling natural enemies, ability to effectively compete in a new ecosystem, availability of artificial or disturbed habitats, and intrinsic adaptability to novel conditions (Hierro et al. 2005; Lloret et al. 2005; Pimentel et al. 2000). As the spread of invasive species becomes increasingly more common globally, it is important to have early detection methods, but distribution maps are not available for many invasive species. This makes it difficult for land managers to focus their efforts to control the spread of invasive species before they become unmanageable. The habitat suitability map produced from this study shows locations that are suitable and likely to be occupied by Giant Hogweed, and such a map can help land managers and policy makers to focus their prevention and control efforts, contain the weed, and finally eradicate it.

References

Albert, R., Christian, S. (2007). *Interactions of Elevation, Aspect, and Slope in Models of Forest Species Composition and Productivity*. Forest Science *53* (4).

Allen, C., Johnson, A., Paris, L. (2006). *A framework for spatial Risk Assessments: Potential impacts of nonindigenous invasive species on native species*. Ecology and Society, *11(1),* 52-66

Almasi, K. (2000). *A non-native perennial invades a native forest*. Biol. Invasions *2,* 219–230.

Andersen, M., Adams, H., Hope, B., Powell, M. (2004a). *Risk analysis for invasive species: General framework and research needs*. Risk Analysis 24-893.

Austin, P. (2002). *Spatial prediction of species distribution: an interface between ecological theory and statistical modelling*. Ecol. Model. *157*, 101–18

Baillie, J., Hilton, T., Stuart, S. (2004). *2004 IUCN Red List of Threatened Species. A Global Species Assessment.* IUCN Gland, Switzerland and Cambridge, UK.

Bethany, A., Mustard, J. (2006). *Characterizing the landscape dynamics of an invasive plant and risk of invation using remote sensing*. Ecological Applications, *16(3)*, 1132-1147.

Brenning, A. (2005). *Spatial prediction models for landslide hazards: review, comparison and evaluation.* Natural Hazards and Earth System Sciences, *5*, 853–862.

Brito, J., Crespo, E., Paulo, O. (1999). *Modeling wildlife distributions: logistic multiple regression vs overlap analysis*. Ecography, *22*, 251-260.

Brothers, T., Spingarn, A. (1992). *Forest fragmentation and alien plant invasion of central Indiana old-growth forests.* Conserv. Biol. *6*, 91–100

Bunkse, E. (2000) *Reality of Rural Landscape Symbolism in the Formation of a Post-Soviet, Postmodem Latvian Identity.* Norsk Geografisk Tidsskrift, *53*, 121 – 138.

Busby, J. (1991). *BIOCLIM—a bioclimatic analysis and prediction system*. Pages 64–68 in C. R. Margules and M. P. Austin, editors. Nature conservation: cost effective biological surveys and data analysis. Commonwealth Scientific and Industrial Research Organization, Melbourne, Victoria, Australia.

Cadanasso, M., Pickett, S. (2001). *Effect of edge structure on the flux of species into forest interiors*. Conserv. Biol. *15*, 91–97

Central Intelligence Agency 2009. *The world fact book 2010*. Potomac Books Inc. Washington, D.C.

Charlotte N., Per, H., Kollmann, J. (2008). *Predicting the didtribution of the invasive alien Heracleum mantegazzianum at two different spatial scales*. Diversity and Distributions *14(2)*, 307-317.

Cramer, J. (2003). *Logit models: from economics and other fields*. Cambridge University Press.

David, A., Rauschert, E., Nord, A., Jones, B. (2009). *Forest roads facilitate the spread of invasive plants.* Invasive Plant Science and management *2*, 191-199

Davis, M., Grime, J., Thompson, K. (2000) *Fluctuating resources in plant communities: a general theory of invisibility*. Journal of Ecology, *88*, 528-534.

Eberhards, G. (1984) *Reljefs (The relief).* In: Latvijas enciklopedija (Encyclopedia of Latvia), *5(2),* 34-41.

Elith, J., Graham, C., Anderson, R., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., Leathwick, F., Lehmann, A., Li, J., Lohmann, L., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., Peterson, A., Phillips, S., Richardson, K., Scachetti-Pereira, R., Schapire, R., Soberon, J., Williams, S., Wisz, M., Zimmermann, N. (2006). *Novel methods improve prediction of species' distributions from occurrence data* Ecography, 129–151.

Elith, J., Leathwick, J. (2009). *Species distribution models: Ecological explanation and prediction across space and time*. Annual Reviews of Ecology, Evolution and Systematics, *4*0, 677–697.

Eurostat. (2010). *Growth rate of real GDP per capital*. European Environment Agency 2010. Land use (Latvia). http://www.eea.europa.eu/soer/countries/lv/soertopic_view?topic=land. Accessed 30 October 2012.

Evans, M., Hastings, N., Peacock, B. (2000). *Bernoulli distribution. Statistical distributions,* 3rd edn. 31–33. Wiley, New York.

Fleishman, E., Nally, M., Fray, J. (2002). *Validation tests of predictive models of butterfly occurrence based on environmental variables.* Conservation biology, *17*, 806-817.

Forman, T., and Alexander, L. (1998). *Roads and their major ecological effects*. Annual Review of Ecology and Systematics *29*, 207-231.

Garcia-Robledo, C., Murcia, C. (2005) *Comparative habitatsusceptibility to invasion by Chinese ash (Fraxinuschinensis: Oleaceae) in a typical Andean landscape*. Biol Invasions *7*, 405–415.

Gavrilova, Ģ. (2003). *Report on investigation of Heracleum in Latvia in 2003*. - Institute of Biology.

Gavrilova, Ģ., Roze, I. (2005). *Species of genus Heracleum L. and their distribution in Latvia.* - 3rd International conference "Research and conservation of biological diversity in Baltic region", Book of abstract. Daugavpils University, Daugavpils, Latvia; 20 – 22 April, 56.

Graham, C., Ferrier, S., Huettman, F., Moritz, C., Peterson, A. (2004). *New developments in museum-based informatics and applications in biodiversity analysis.* Trends in Ecology and Evolution, *19*, 497–503.

Greenburg, C., Crownover, S., Gordon, D. (1997). *Roadside soils: a corridor for invasion of xeric scrub by nonindigenous plants.* Natural Areas Journal *17(2)*, 99-109.

Guisan, A., Zimmermann, N. (2000). *Predictive habitat distribution models in ecology*. Ecol. Modell. *135*, 147–186.

Harrison, S. (1999). *Local and regional diversity in a patchy landscape*: native, alien, and endemic herbs on serpentine. Ecology *80*:70-80.

Harrison, S., Hohn, C., Ratay, S. (2002). *Distribution of exotic plants along roads in a peninsular nature reserve.* Biol. Invasions *4*, 425–430.

Hernandez, P., Graham, C., Master, L., Alber, D., Hernandez, P., et al. (2006). *The effect of sample size and species characteristics on performance of different species distribution modeling methods* Ecography, *29*, 773–785.

Hierro, J., Maron, J., Callaway, R. (2005). *A biogeographic approach to plant invasions: the importance of study of exotics in their introduced and native range*. J Ecol *93*, 5–15.

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

Hobbs, R. (1989). *The nature and effects of disturbance relative to invasion.* In: Drake J, Mooney H, Dicastri F, Groves R, Kruger F, Rejmanek M, Williamsson M (eds), Biological invasions: a global perspective, Scope. Wiley, New York, *37*, 389–405.

Hobbs, R., Huenneke, L. (1992) *Disturbance, diversity, and invasion: implications for conservation.* Conserv Biol *6*, 324–337.

Hosmer, D., Lemeshow, S. (2000). *Applied logistic regression*, 2nd edition. John Wiley & Sons, Inc., New York, New York.

Huntley, B., Berry, P., Cramer, W., Mcdonald, A. (1995). *Modelling present and potential future ranges of some european higher plants using climate response surfaces*. Journal of Biogeography, *22*, 967-1001.

JFNew. (2009). website: http://www.jfnew.com/species-control.asp. Accessed 22 December 2009.

Kabuce N. (2006): NOBANIS – *Invasive Alien Species Fact Sheet – Heracleum sosnowskyi.* – From: Online Database of the North European and Baltic Network on Invasive Alien Species – NOBANIS www.nobanis.org.

Kabuce, N., Priede, N. (2010): NOBANIS – *Invasive Alien Species Fact Sheet – Heracleum sosnowskyi.* – From: Online Database of the European Network on Invasive Alien Species – NOBANIS www.nobanis.org. Access 10 June 2010.

Kadoya, T., Washitani, I. (2007). *An adaptive management scheme for wetland restoration incorporating participatory monitoring into scientific predictions using dragonflies as an indicator Taxon.* Global Environmental Research *11*, 179-185.

Knight, K., Reich, P. (2005). *Opposite relationships between invisibility and native species richness at patch versus landscape scales.* Oikos *109*, 81–88.

Kumar, S., Stohlgren, T., Chong, G. (2006). *Spatial heterogeneity influences native and nonnative plant species richness.* Ecology, *87*, 3186-3199.

Laivins, M., Gavrilova, G. (2003) *Heracleum sosnowskyi in Latvia: sociology, ecology and distribution.* Latvijas Veģetācija ***7***, 45-65.

Lambrinos, J. (2000). *The variable invasive success of Cortaderia species in a complex landscape.* Ecology, *83*, 518-529.

Larson, D., Anderson, P., Newton, W. (2001). *Alien plant invasion in mixed-grass prairie: Effects of vegetation type and anthropogenic disturbance*. Ecological Applications *11(1),* 128-141.

Leathwick, J., Elith J., Francis, M., Hastie, T., Taylor, P. (2006) *Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees.* Mar Ecol Prog Ser *321*, 267–281.

Leathwick, J., Elith, J., Hastie, T., 2006. *Comparative performance of generalized additive models and multivariate adaptative regression splines for statistical modelling of species distributions.* Ecol. Model. *199*, 188–196.

Lehmann, A., Overton, J., Austin, M. (2002). *Regression models for spatial prediction: their role for biodiversity and conservation.* Biodiversity and Conservation, *11*, 2085–2092

Levine, J. (2000). *Species diversity and biological invasions: relating local process to community pattern.* Science *288*, 852-854

Liu, C., White, M., Newell, G., Griffioen, P. (2011) *Species distribution modelling for conservation planning in Victoria, Australia*. MODSIM 2011. Perth, Western Australia

Lloret, F., Me´ dail, F., Brundu, G., Camarda, I., Moragues, E., Rita, J., Lambdon, P., Hulme, P. (2005). *Species attributes and invasion success by alien plants on Mediterranean.* J.Ecol. *93*, 512-520

Mack, R., Simberloff, D., Lonsdale, W., Evans, H., Clout, M., Bazzazz, F. (2000) *Biotic invasions: causes, epidemiology, global consequences, and control*. Ecol Appl *10*, 689–710

Manel, S., Dias, J., Ormerod, S. (1999). *Comparing discriminant analysis, neural networks and logistic regression for prediction species distributions: a case study with a Himalayan river bird.* Ecological Modelling, *120*, 337-347.

McCullagh, P., Nelder, J. (1989). *Generalized Linear Models*, Second Edition. Boca Raton: Chapman and Hall/CRC. ISBN 0-412-31760-5.

NASA Science – *Earth education website for invasive species*, http://develop.larc.nasa.gov/about/natpriority.html. Updated January 15, 2009. Accessed August 17, 2009.

Newbold, T. (2010). *Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models*. Progress in Physical Geography, *34*, 3-22.

Nielsen, C., Ravn, H., Nentwig, W., Wade, M. (2005). *The Giant Hogweed Best Practice Manual. Guidelines for the management and control of an invasive weed in Europe*. Forest & Landscape Denmark, Hoersholm, *44*.

Normunds, P. (1993). *Black Alder Swamps on Forested Peatlands in Latvia* Folia Geobotanica & Phytotaxonomica, *28(3)* 261-277 Published by: Springer.

Obolevicˇa, D. (2001). *Hogweed and its distribution in Latvia*. http:// www.lva.gov.lv/daba/eng/biodiv/lauks_latvanis_e.htm.

Orians, G. (1986). *Site characteristics favoring invasions* In: Mooney H, Drake J (eds) Ecology of biological invasions of North America. Springer-Verlag, New York, 113–148.

Osborne, P., Alonso, J., Bryant R. (2001). *Modeling landscape-scale habitat-use using GIS and remote sensing: a case study with great bustards*. J. Appl. Ecol., *38*, 458-471.

Ota. (1993). *Harmful Non-indigenous Species in the United States*. U.S. Government Printing Ofiice, Washington, D.C. 391.

Pearson, R., Dawson, T., Berry, P., Harrison, P. (2002). Species: *A spatial evaluation climate impact on the envelop of species.* Ecological modeling *154*, 289-300.

Phillips, S., Dudik, M. (2008). *Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation* Ecography, *31* (2008), 161–175.

Pimentel, D., Lach, L., Zuniga, R., Morrison, D. (2000). *Environmental and economic costs associated with non-indigenous species in the United States*. BioScience *50*, 53-65.

Planty-Tabachhi, E., Tabachhi, E., Nieman, R., DeFerrari, C., Decamps, H. (1996). *Invasibility of species rich communities in riparian zones*. Conserv Biol *10*, 598–607.

Pulliam, H. (2000). *On the relationship between niche and distribution*. *Ecology Letters*, *3*, 349–361.

Pyšek, P. (1998). *Alien and native species in Central European urban floras: a quantitative comparison*. – J. Biogeogr. *25*, 155–163.

Pysek, P., Jarosik, V., Mullerov, J., Pergl, J., Wild, J. (2007). *Comparing the Rate of Invasion by Heracleum mantegazzianum at Continental, Regional and Local Scales, Diversity and Distribution, 14*, 355-363.

Rand, T., Russell, F., Louda, S. (2004). *Local-vs. landscape-scale indirect effects of an invasive weed on native plants.* Weed Technol *18*, 1250–1254.

Robertson, M., Peter, C., Villet, M., Ripley, B. (2003) *Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modeling techniques.* Ecological Modeling, *164*, 153–167.

Silveri, A., Dunwiddie, P., Michaels, H. (2001). *Logging and edaphic factors in the invasion of an Asian woody vine in a mesic North American forest*. Biol. Invasions *3*, 379–389.

Smith, P. (1994). *Autocorrelation in logistic regression modeling of species distributions*. Global Ecology and Biogeography Letters, *4*, 47-61.

Stockwell, D. (2006). *Improving ecological niche models by data mining large environmental datasets for surrogate models*. Ecological modeling *192*, 188-196

Stockwell, D., Peters, D. (1999). *The GARP modelling system: problems and solutions to automated spatial prediction.* Int. J. Geogr. Inf. Sci. *13*, 143–58

Stohlgren, T., Barnett, D., Kartez, J. (2003). *The rich get richer: patterns of plant invasions in the United States.* Frontiers in Ecology and the Environment, *1*, 11-14.

Stohlgren, T., Lee, M., Bull, K., Otsuki, Y., Villa, C. (1998). *Riparian zones as havens for exotic plant species in the central grasslands.* Plant Ecology *138(1),* 113-125

Stohlgren, T., Ma, P., Kumar, S., Rocca, M., Morisette, J., Jarnevich, C., Benson, N. (2010). *Ensemble Habitat Mapping of Invasive Plant Species*. Risk Analysis, *30(2)*, 2010

Stohlgren, T., Schell, L., Vanden, B. (1999b). *How grazing and soil quality affect native and exotic plant diversity in rocky mountain grasslands*. Ecological Applications *9(1)*, 45-64

Store, R., Kangas, J. (2001). *Integrating spatial multi-criteria evaluation and expert knowledge for GIS-based habitat suitability modeling*. Landscape and Urban Planning *55*, 79-93

Taff, G., Müller, D., Kuemmerle, T., Ozdenerol, E., Walsh, S. (2010) *Reforestation in Central and Eastern Europe After the Breakdown of Socialism*. In: Nagendra H, Southworth J, (eds.), Reforesting Landscapes: Linking Pattern and Process, Springer Netherlands (Landscape Series, *10*, 121 – 147)

Thomas, J., Middleton, B., Gibson, D. (2006). *A landscape perspective of the stream corridor invasion and habitat characteristics of an exotic (Dioscorea oppositifolia) in a pristine watershed in Illinois.* Biol Invasions *8*, 1103–1111

Trombulak, S., Frissell, C. (2000). *Review of ecological effects of roads on terrestrial and aquatic communities.* Conservation Biology *14*, 18-30

Tyser, R., Worley, C. (1992). *Alien flora in grasslands adjacent to road and trail corridors in Glacier National Park, Montana (U.S.A.).* Conservation Biology *6(2)*, 253-262

U.S. Fish and Wildlife Service, (2003). *Costly fires hurt wildlife habitat*. [Online] http://www.fws.gov/deerflat/currfire.htm

Vitousek, P., D'Antonio, c., Loope, L., Westbrooks, R. (1996). *Biological invasions as global environmental change.* American Scientist *84*, 468-478

Wilcove, D., Rothstein, D., Dubow, J., Phillips, A., Losos, E. (1998). *Quantifying the threats to imperiled species in the United States*. BioScience *4*8, 607–615

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

Chapter 4

Ecological niche modeling using cluster analysis to determine suitable environments for Giant Hogweed presence

Introduction

Invasive species drive ecological dynamics at multiple spatial scales and levels of organization, through local and regional extinctions of native species (Mack et al. 2000) and entire communities, shifts in native species richness and abundance (Parker et al. 1999), and altered fire regimes, water quality, and biogeochemical cycles (D'Antonio and Vitousek 1992; Vitousek et al. 1996; Strayer et al. 1999; Bohlen et al. 2004). Invasive species are the second leading cause (after human population growth and associated activities) of species extinction and endangerment in the US (Pimentel 2002). The most common strategy for estimating the actual or potential geographic distribution of a species is to characterize the environmental conditions that are suitable for the species, and to then identify where suitable environments are distributed in space (Pearson et al. 2007). Some types of land use can exacerbate the spread and effects of invasive species across scales (Dukes and Mooney 1999; Simberloff 2000). Identifying invasion and curtailing the spread of invaders is an enormous ecological and societal challenge (Lodge et al. 2006). Many plant populations are declining and face increasing threat from human disturbances and the proliferation of nonindigenous species. For example, fragmentation of habitat may increase the distribution and abundance of many invasive plants (With 2002, 2004).

Invasive alien plants such as Giant Hogweed give increasing cause for concern. Giant Hogweed was introduced to Latvia from the Caucuses in 1948 as a silage plant for livestock because of its hardiness and large biomass (Pysek et al. 2007). During this time, little was known about the toxicity of the plant, and it has since then spread out to form dense patches in many regions of the country, threatening biodiversity, ecological functioning and human health. According to a 2001 survey, Giant Hogweed had invaded and occupied over 12,000 hectares of land in Latvia (Obolevica 2008). The plant is poisonous to humans and domesticated animals and can be fatal if ingested (Benezra 1989). Local residents are afraid and have very special concerns about the safety of their children because the plant causes phytophotodermatitis (severe burns), which results in painful blistering, permanent scarring when the sap of the plant comes in contact with human skin and is exposed to sunlight, and blindness when the sap comes in contact with eyes. Giant Hogweed also has severe negative impacts on a variety of ecosystems leading to a reduction in local plant biodiversity, considerable economic loss and health hazards to humans (Nielsen 2005). No comprehensive tool exists to stop these invasive plants, reduce their impact or prevent future invasions. Giant Hogweed is present throughout the whole country of Latvia and is spread via natural and human means. Therefore, landscape-level effects (topographic and distance variables) should be important, but should be studied on a regional level since human drivers of species spread may vary throughout different parts of the country.

Spatial modeling is a promising approach to predicting risk of invasion because spatial patterns of invasion can be predicted by linking current presence and absence of invasive species to spatially explicit predictor variables, such as land use, geomorphology, and topography, using geographic information systems (GIS); (Store and Kangas 2001). Applying a predictive distribution model within a spatial context relies on the existence of landscape-scale variables that define suitable habitat for a species (Osborne, Alonso and Bryant 2001; Austin 2002a, b) based on the biological/ecological needs of that species.

Cluster analysis is a class of techniques used to classify cases into groups that are relatively homogeneous among themselves and heterogeneous between each other, on the basis of a defined set of variables. These groups are called clusters.

Cluster analysis was proposed in previous studies as a useful tool for the selection of sites with representative environmental conditions (Mackey et al. 1988; Belbin 1993, 1995; Kirkpatrick and Brown 1994; Faith and Walker 1996a; Fairbanks and Benn 2000). The objective of cluster analysis is ascription of the objects in question into groups (clusters), so as to maximize the similarity between the members of each group and to minimize the similarity between groups (Legendre and Legendre 1998). Hence, sites could be classified using cluster analysis into relatively homogenous groups, different from each other, on the basis of their values of environmental factors. In this study cluster analysis was used to group the environmental factors that promote Giant Hogweed presence. We therefore hypothesize that cluster analysis can be used to accurately predict sets of environmental variable values that characterize where Giant Hogweed thrives, and will likely grow in the future based on a set of environmental conditions, e.g. in low flat areas near roads.

The first main objective of this study is to use cluster analysis to group the most common classes of conditions where Giant Hogweed is present to understand its key habitats in this landscape. For example Giant hogweed may generally not flourish in high elevation habitats but if there is development such as roads construction or buildings on high elevation habitats they may become suitable for Giant hogweed to grow. The second main objective of this study is to determine what sets of environmental and land use conditions characterize each of the clusters, and 3) use the results from cluster analysis to create a Giant Hogweed suitability map which can be used to determine priority areas for conservation planning.

To validate the performance of the model the data was divided into training and test data sets using random sampling in SPSS. The training data (70%) was used for training and calibration of the model while the test data (30%) was used to test the performance of the model.

Methods

Study area

The study was carried out in northeast Latvia which lies on the Baltic coast, in the western part of Eastern Europe centered between 57’ 96” and 56’ 03” N-latitude and 25’ 35” and 27’ 45” E-longitude. Latvia is one of the three Baltic States, situated on the east side of the Baltic Sea, the others being Estonia (to the north) and Lithuania (to the south). Latvia also borders Russia and Belarus to the east (fig. 1). The total land area is 64.6 thousand km^2 and the terrain is mostly low plain, with majority of the territory between 40-200 meters above sea level (Eberhards 1984). The climate is wet with moderate winters for this latitude. The average amount of precipitation is 600 – 650 mm annually; the vegetation period usually last for 180 – 200 days (Normunds Prieditis 1993). The landscape is characterized by mature forests, secondary forests, meadows, farmland, abandoned farmlands, lakes, rivers, hills, plains, villages and dispersed rural homesteads (Bunkse 2000). Some of these land cover types have been invaded by Giant Hogweed where it develops large stands and dominates native species in occupied territories. It is easy to distinguish it from the surrounding vegetation especially during summer when the plant flowers. This region is chosen as the study site because it is where the species was first introduced in the country and existing records of the study species in this region show a substantial distribution, and many Giant Hogweed patches are large when compared to other areas.

Environmental variables used for the cluster analysis

Three types of environmental variables were used in the cluster analysis, including topographic, land-cover land-use change, and demographic data. These landscape characteristics often play an important role in species distribution and patterns of invasion (lambrinos 2002; Stohlgren et al. 2003; Kumar et al. 2006) and have been used extensively in other studies as factors affecting invasive species distribution (Bethany et al. 2006; Almasi 2000; Silveri et al 2001). Thirteen environmental variables were used: elevation, slope, neighborhood land cover change (no change, change to agriculture, change to forest, and change to urban), neighborhood land cover types (forest cover, agriculture cover, urban cover), distance to roads, distance to water, distance to urban centers, and human population density. LANDSAT data was used for the land cover data which were obtained from the USGS Visualization Viewer (Glovis), demographic data were obtained from the Latvian Bureau of Statistics, DEM and other GIS layer were obtained from free and open source GIS to make maps of species distribution data (DIVA GIS). Giant Hogweed presence location points were obtained from Agnese Priede's research and from the Latvia Ministry of Agriculture, and supplemented with the authors' fieldwork and public participation geographic Information systems (PPGIS) organized by our research group at the University of Memphis.

Fig. 1 Study area - northeastern Latvia

Species Modeling

Several ecological niche models such as bioclimatic prediction system (BIOCLIM) (Busby 1991), generic algorithm for rule-set prediction modeling system (GARP) (Stockwell and Peters 1999; Stockwell et al. 2006), artificial neutral network (ANN) (Pearson et al. 2002), ecological niche factor analysis (ENFA) (Hirzel et al. 2002) and regression methods such as generalized linear models (GLM) (Lehman et al. 2002), generalized additive models (GAM) (Elith et al. 2006), boosted regression trees (BRT) (Leathwick et al. 2006) and multivariate adaptive regression splines (MARS) (Elith et al. 2007) have been widely used to predict potential geographic distribution of non-native species. Cluster analysis was selected for this study because it is a novel modeling technique that can uses presence-only data and uniquely gives a combination of environmental factors that promote the presence of invasive species. Although predictive models are often, and perhaps best, built using techniques such as logistic regression that rely on both presence and absence data (e.g., Peeters and Gardeniers 1998; Manel et al. 1999; Mladenoff et al. 1999), in many applications absence data are unavailable, unreliable, or incomplete. At the time of a survey, a species may have undergone declines for reasons unrelated to habitat quality (van Manen et al. 2005; Thompson et al. 2006). On the other hand, many predictive models been developed to predict distribution or rank potential habitat using only presence data (e.g., Busby 1991; Clark et al. 1993; Hirzel et al. 2002; Lele and Keim 2006). Cluster analysis in this study uses presence only data. One disadvantage of presence only models is that presence-only locations are not guaranteed to be unbiased with respect to the spatial distribution of the species. In this research sampling was biased towards roads because of the serious health issues such as severe burns caused by Giant Hogweed which can lead to bias prediction in species occurrence. Nonetheless presence only data have been used widely to predict species distribution. There is certainly more work to do in developing statistical models for the analysis of presence-only data that account for sample selection bias and for errors in detection of species.

Clustering methods

Cluster analysis is a class of techniques used to classify cases into groups that are relatively homogeneous within themselves and heterogeneous between each other, on the basis of a defined set of variables. These groups are called clusters. There are numerous ways in which clusters can be formed. Hierarchical clustering is one of the most straightforward methods. It can be either agglomerative (building clusters from individuals) or divisive (starting from 1 cluster, breaking it into *n* cluster). Agglomerative hierarchical clustering begins with every case being a cluster unto itself. At successive steps, similar clusters are merged. The algorithm ends with cases in a smaller set of clusters. In agglomerative clustering, once a cluster is formed, it cannot be split; it can only be combined with other clusters. Agglomerative hierarchical clustering is such a method, which doesn't let cases separate from clusters that they've joined: once in a cluster, always in that cluster. If variables are measured on different scales, variables with large values contribute more to the distance measure than variables with small values. If variables are measured on the same scale (through standardizing variables, for instance), then this not a problem. Hierarchical clustering chooses the number of final clusters, and write in basic terms which parameters or outcomes this decision can be based on. It computes all possible distances and specify distances between each case and all clusters.

A clustering method that does not require computation of all possible distances is *k*-means clustering. It differs from hierarchical clustering in several ways. One has to know in advance the number of clusters one wants. First assignment of the k cluster means is arbitrary. The algorithm is called *k*-means, where *k* is the number of clusters one wants, since a case is assigned to the cluster for which its distance to the cluster mean is the smallest. After all cases are assigned, new cluster means are calculated. This step is repeated until cluster means don't change much between successive steps (based on a predefined threshold). Finally, the algorithm calculates the means of the clusters once again and assigns the cases to their permanent clusters. *K*-means clustering is very sensitive to outliers.

Data Processing

There were a total of 477 cases and data were divided into training and test datasets. 75% of the data was used to train the model and 25% was reserved for testing. Euclidean Distance was calculated as distance to the nearest road, river and urban center. Slope was derived from a Digital Elevation Model using the Spatial Analysis tool in ArcMap 10. New raster data layers were created using 5x5 neighborhood filters created for each of the land cover types and land cover change classes, counting the total number of each class represented within the 5x5 filter. These values range from 0 to 25 (table 1). Giant Hogweed location points were then overlaid onto the land cover and LCC maps and the values of these raster grids were extracted to each Giant Hogweed point. The result of this process was to provide a characterization of land cover and LCC in neighborhoods (5x5 30m-cell neighborhoods) around each Giant Hogweed point.

In this study, agglomerative hierarchical clustering was performed using Ward's method applying the Euclidean Distance as the distance or similarity measure to define the number of clusters. Agglomerative cluster was chosen because it uses a bottom-up algorithm that treats each case as single cluster and then successfully merged (or agglomerate) pairs of clusters until all clusters have been merged into a single cluster that contain all the cases. This method assumes that the merged operation is monotonic. This method has the advantage it can produce an ordering of objects which may be informative for data display. The method is also simple and you do not need to specify any number of clusters from the beginning. The main disadvantage of this method is that no provisions can be made for a relocation of cases that may have been incorrectly grouped from the beginning.

All variables were first standardized using Z-scores and outliers deleted. The number of clusters was determined from the agglomeration schedule table which provides a solution for every possible number of cluster like in this study from 1-377 (i.e. identify the step where the “distance coefficient” number makes a big jump and then subtract this number from the total number of cases): 377 – 373 = 4 (Kaufman et al., 2005). After the number of clusters (four) was determined using agglomerative hierarchical clustering, a k-Means cluster was run with the specification that four clusters should be formed. The descriptive statistics for all (unstandardized) variables for the final clusters, including cluster centers (means), are shown (tables 4). The clusters were evaluated using F ratios in table 5. F ratios describe the differences between the clusters and can be tested for statistical significance.

Results

Table 1 Descriptive statistics for the unstandardized independent variables

	N	Minimum	Maximum	Mean	Std. Deviation
Elevation	377	39	245	109.58	42.689
Distance to road (km)	377	.00	3.34	.4541	.56458
Distance to urbancenter (km)	377	.00	36.85	11.2188	7.51846
Distance to water (km)	377	.03	12.30	3.7393	2.20015
Slope	377	.00	18.43	.7311	2.22460
No change	377	0	25	21.64	4.480
Forest change	377	0	25	1.18	3.054
Agriculture change	377	0	22	1.05	2.391
Urban change	377	0	15	1.13	2.492
Urban land cover	377	0	23	1.82	3.288
Agriculture land cover	377	0	25	20.07	5.980
Forest land cover	377	0	25	2.55	5.303
Population density	377	0	133	12.40	11.322
Valid N (listwise)	377				

Results of Hierarchical-cluster analysis

Agglomeration schedule

In the column labelled *Coefficients*, the value of the distance (or similarity) statistic used to form clusters is found. From these numbers, you get an idea of how unlike the clusters being combined are. These coefficients help to decide how many clusters to represent the data. The distance coefficient makes the largest jump from 7.108 to 9.326 between stages 373 and 374. Therefore, based on criterion above, the ideal number of clusters = 377-373=4

Table 2 Defining the number of clusters using portion of the agglomeration schedule table. The columns labeled *Stage Cluster First Appears* tell you the step at which each of the two clusters that are being joined first appear e.g. at stage 371, cluster 1 combines with cluster 136, the resulting cluster is cluster 1 and that cluster 1 will still action in the *next stage* 372

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
367	136	193	5.399	351	364	371
368	10	377	5.442	363	0	369
369	10	303	5.776	368	353	372
370	1	191	5.830	360	365	371
371	1	136	6.018	370	367	372
372	1	10	6.471	371	369	373
373	1	125	7.108	372	366	374
374	1	240	9.326	373	362	375
375	1	239	10.554	374	0	376
376	1	327	11.800	375	0	0

Results of K-cluster analysis

Table 3 Standardized variables for the final cluster centers.

	Standardized Cluster Means			
	1	2	3	4
Zscore(agricultural change)	-.24007	.90767	-.07773	-.16594
Zscore(elevation)	.08065	-.05849	-.33569	-.07463
Zscore(distance to road km)	.03159	-.42826	.33472	-.25259
Zscore(distance to urban center km)	.21010	-.43524	-.45707	-.55315
Zscore(distance to water km)	.04965	-.11011	-.03520	-.19479
Zscore(forest change)	-.23607	-.31313	1.91316	-.23138
Zscore(agriculture land cover)	.48235	-.67534	-1.93205	.12605
Zscore(forest land cover)	-.32168	-.29858	2.35100	-.12066
Zscore(urban land cover)	-.34958	1.78076	-.44222	.01226
Zscore(nochange)	.47631	-1.24443	-1.09161	.27748
Zscore(population density)	-.15205	.08813	.61235	.49152
Zscore(slope)	-.23951	-.20978	-.18295	3.12492
Zscore(urban change)	-.34824	1.84792	-.36147	-.05113

The final cluster centers are formed when all the cases have been assigned to a cluster and the cluster centers are computed one last time. The final cluster centers are used to describe the clusters by variables (discussion section).

Table 4 Distances (in standardized variable space) between final cluster centers

Cluster	1	2	3	4
1		3.965	4.625	3.593
2	3.965		5.031	4.714
3	4.625	5.031		5.342
4	3.593	4.714	5.342	

Table 4, shows the Euclidean distance between the final cluster centers. Greater distances between clusters mean there are greater dissimilarities e.g. cluster 3 and cluster 4 have the greatest dissimilarity because the distance between them is 5.342, which is greater than the distance value between any other two clusters.

Cluster membership

Cluster membership shows the number of cases in each cluster. Fig. 2, shows the number of cases in cluster 1, cluster 2, cluster 3 and cluster 4. A majority of the cases are in cluster 1 (252) and the least number of cases are in cluster 4 (26). This simply means that cluster 1 represents the most common set of conditions where Giant Hogweed is found in the Latvian landscape in this part of the country. This most common cluster is represented by agricultural lands where little land-use change has occurred in the neighbourhood (see detailed cluster descriptions below in Discussion section).

Table 5 Number of cases in each cluster

Cluster	1	252.000
	2	60.000
	3	39.000
	4	26.000
Valid		377.000
Missing		.000

F-ratios table

F-ratios are shown in the ANOVA table (Table 5) which presents the F values and significance levels to show which mean differences are significant. Most of the between groups means are significant, indicating eleven variables reliably distinguish between the four clusters. The significance level of distance to water and elevation are >0.05 meaning they did not contribute much to the separation of the clusters (table 5)

Table 6 F-test and significance of variables in the model

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Zscore(agriculture change)	21.477	3	.771	373	27.855	.000
Zscore(elevation)	2.125	3	.971	373	2.189	.089
Zscore(distance to road inkm)	5.650	3	.914	373	6.178	.000
Zscore(distance tourbancenter inkm)	12.839	3	.894	373	14.358	.000
Zscore(distance to water inkm)	.794	3	1.055	373	.753	.521
Zscore(forest change)	54.606	3	.526	373	103.717	.000
Zscore(land-cover agriculture)	77.258	3	.366	373	211.123	.000
Zscore(land-cover forest)	82.359	3	.288	373	285.990	.000
Zscore(land-cover urban)	76.228	3	.441	373	172.950	.000
Zscore(no change)	66.099	3	.461	373	143.285	.000
Zscore(populationden sity)	9.054	3	1.031	373	8.785	.000
Zscore(slope)	90.764	3	.330	373	274.725	.000
Zscore(urban change)	80.152	3	.484	373	165.553	.000

Habitat suitability maps for each cluster

Habitat suitability maps may be created using binary habitat suitability maps or ranking suitability maps. Binary suitability maps classify landscape as either good or bad, with no other choice. Ranked suitability classifies the landscape using a range of values from bad to good. Binary suitability is simple but has the disadvantage of no "in-between" choices. To create a habitat suitability map for each cluster, areas were demarcated throughout the study site whose variable values fit those of each cluster. To do so, the 10^{th} and 90^{th} percentiles for each variable were calculated for the cases within each cluster, and all values of variables between these percentiles therefore contains the majority of the cases in the center range of each variable within the cluster. Ranked habitat suitability was then used to classify all the different clusters. In this case, the number of variables for which each pixel fell in the appropriate range (between the 10^{th} percentile and 90^{th} percentile) was summarized for each cluster. Since there were 13 variables, each pixel received a rank from 0 to 13, with 13 being the highest suitability for that cluster (map shown for Cluster 1 in Figure 2). A map of the most suitable areas for Cluster 1 is shown in Figure 3 (showing areas where between 10 and 13 variables in the 10^{th} – 90^{th} percentile range overlap). The main disadvantage of this method is that it cannot determine what factors contributed to the final value. The results of these analyses show how suitable each pixel in the landscape is to be colonized by Giant Hogweed in the future, based on any of the 4 clusters (Figure 4). The percentiles for each cluster are shown in Tables 7, 8, 9 and 10.

Cluster 1 Results

Table 7 Percentile for cluster 1 (10^{th} and 90^{th} percentiles used as endpoints to find likely candidate pixels where Giant Hogweed is currently absent but likely to be present in the future).

Percentiles

		Percentiles						
		5	**10**	25	50	75	**90**	95
Weighted Average(Definition 1)	elevation	61.00	**64.00**	91.00	97.50	141.75	**183.00**	188.00
	Distance to road (km)	.0123	**.0319**	.0862	.2474	.6541	**1.3900**	1.8394
	Distance to urbancenter (km)	2.1430	**4.0578**	6.6659	13.8400	18.3443	**20.5432**	23.3712
	Distance to water (km)	.3375	**1.0211**	2.2960	3.7188	5.1881	**6.3079**	7.8597
	slope	.0000	**.0000**	.0000	.0000	.0000	**.8949**	1.9092
	No change	19.00	**21.00**	23.00	25.00	25.00	**25.00**	25.00
	Forest change	.00	**.00**	.00	.00	.00	**2.70**	3.00
	Agriculture change	.00	**.00**	.00	.00	.00	**2.00**	4.00
	Urban change	.00	**.00**	.00	.00	.00	**1.00**	2.00
	Land-cover urban	.00	**.00**	.00	.00	1.00	**3.00**	4.00
	Land-cover agriculture	17.00	**19.00**	21.00	24.00	25.00	**25.00**	25.00
	Land-cover forest	.00	**.00**	.00	.00	1.00	**4.00**	5.00
	populationdensity	3.00	**5.00**	8.00	8.00	11.00	**17.00**	19.05

Cluster 2 Results

Table 8 Percentile for cluster 2

		Percentiles						
		5	**10**	25	50	75	**90**	95
Weighted Average(Definition 1)	elevation	45.45	**56.00**	77.25	101.00	126.00	**160.00**	190.60
	Distance to road km	.0077	**.0156**	.0371	.1082	.2656	**.4813**	1.1346
	Distance to urban center (km)	.0000	**.0000**	.0893	5.2018	15.6288	**19.8358**	28.2383
	Distance to water (km)	**.3848**	.6264	1.5470	3.2786	5.0505	**6.5010**	7.4694
	slope	**.0000**	.0000	.0000	.0000	.0000	**.0000**	1.8614
	No change	**10.00**	10.00	13.00	16.00	19.75	**21.00**	21.00
	Forest change	**.00**	.00	.00	.00	.00	**1.00**	1.00
	Agriculture change	**.00**	.00	.00	2.00	5.00	**9.80**	13.85
	Urban change	**.00**	1.00	3.00	5.00	8.00	**11.90**	12.95
	Land-cover urban	**.05**	2.00	5.00	7.00	10.00	**13.00**	16.90
	Land-cover agriculture	**7.05**	11.10	13.25	17.00	18.00	**20.90**	22.00
	Land-cover forest	**.00**	.00	.00	.00	1.00	**4.00**	6.00
	Population density	**.00**	5.10	8.00	11.00	17.00	**24.30**	33.00

Cluster 3 Results

Table 9 Percentile for cluster 3

		Percentiles						
		5	**10**	25	50	75	**90**	95
Weighted Average(Definition 1)	elevation	47.00	**50.00**	57.00	92.00	104.00	**171.00**	192.00
	Distance to road (km)	.0355	**.0624**	.1628	.4935	1.0696	**1.4116**	1.6781
	Distance to urban center (km)	.1038	**.6679**	3.4737	6.3964	13.0528	**16.3707**	19.5405
	Distance to water (km)	.2579	**.4398**	1.8361	3.9562	5.4066	**6.5562**	6.8095
	slope	.0000	**.0000**	.0000	.0000	.0000	**1.3502**	2.8624
	No change	5.00	**7.00**	13.00	18.00	20.00	**24.00**	25.00
	Forest change	.00	**.00**	2.00	6.00	12.00	**18.00**	20.00
	Agriculture change	.00	**.00**	.00	.00	1.00	**5.00**	5.00
	Urban change	.00	**.00**	.00	.00	.00	**1.00**	3.00
	Land-cover urban	.00	**.00**	.00	.00	.00	**1.00**	4.00
	Land-cover agriculture	.00	**.00**	3.00	8.00	14.00	**17.00**	19.00
	Land-cover forest	6.00	**8.00**	10.00	15.00	22.00	**25.00**	25.00
	Population density	6.00	**8.00**	8.00	17.00	21.00	**48.00**	48.00

Cluster 4 Results

Table 10 Cluster 4 percentile

		Percentiles						
		5	**10**	25	50	75	**90**	95
Weighted Average(Definition 1)	elevation	44.00	**46.80**	57.75	108.00	139.50	**172.50**	212.90
	Distance to road (km)	.0029	**.0062**	.0252	.2114	.4792	**.9198**	1.2622
	Distance to urban center (km)	.0000	**.0000**	1.7839	5.3849	11.4314	**18.9248**	21.6264
	Distance to water (km)	.7916	**1.2615**	2.2672	3.5314	4.4938	**5.1131**	5.3719
	slope	3.8141	**3.8141**	4.7636	5.7106	9.6940	**12.4978**	16.5796
	No change	16.05	**18.70**	20.75	24.00	25.00	**25.00**	25.00
	Forest change	.00	**.00**	.00	.00	.00	**4.00**	4.65
	Agriculture change	.00	**.00**	.00	.00	1.25	**3.00**	3.00
	Urban change	.00	**.00**	.00	.00	2.00	**3.00**	4.30
	Land-cover urban	.00	**.00**	.00	1.50	4.00	**4.30**	5.00
	Land-cover agriculture	4.55	**13.70**	19.50	22.00	24.25	**25.00**	25.00
	Land-cover forest	.00	**.00**	.00	.00	1.00	**6.30**	18.70
	Population density	4.05	**7.40**	8.75	12.00	18.00	**48.00**	48.00

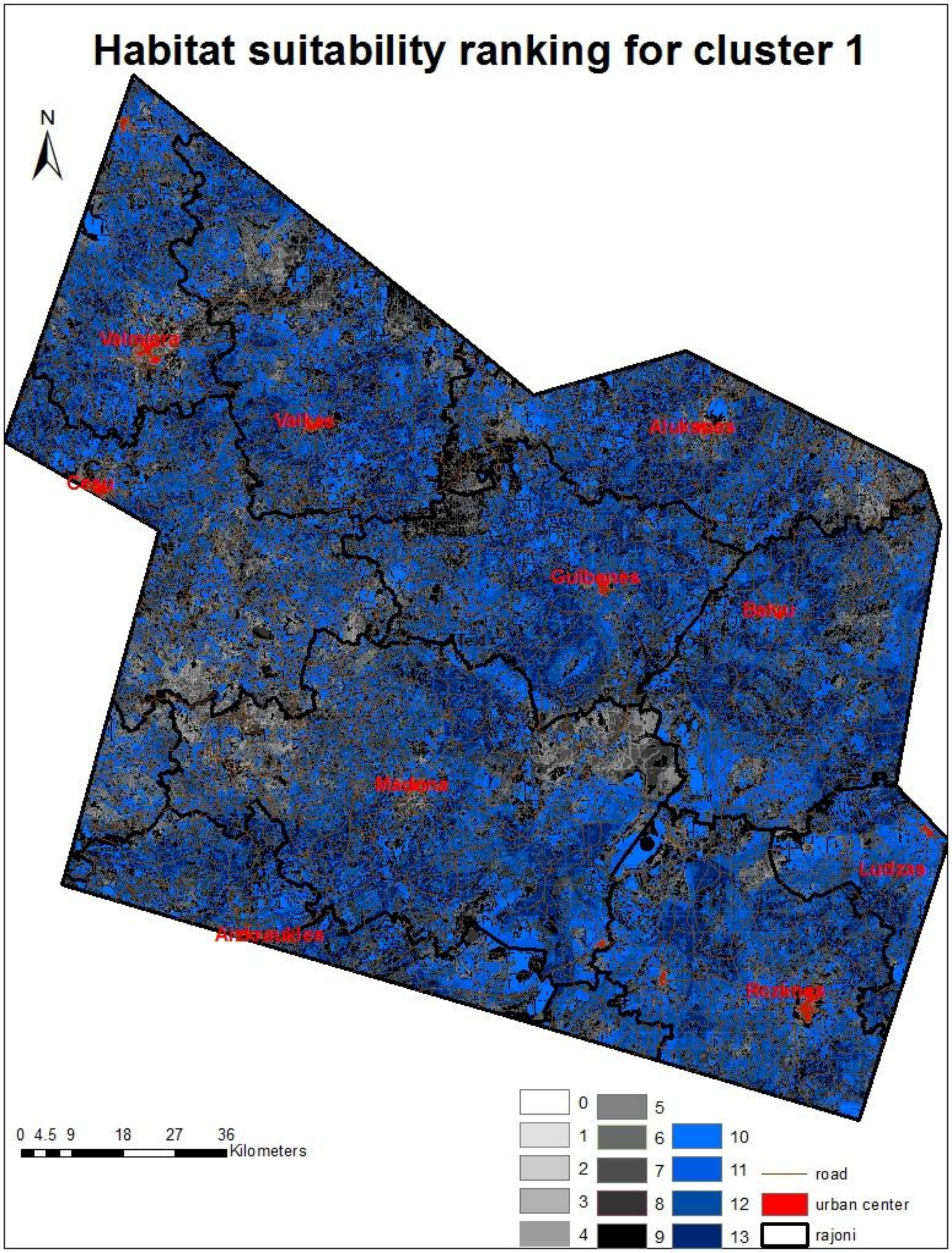

Fig. 2 Habitat suitability ranking for cluster 1

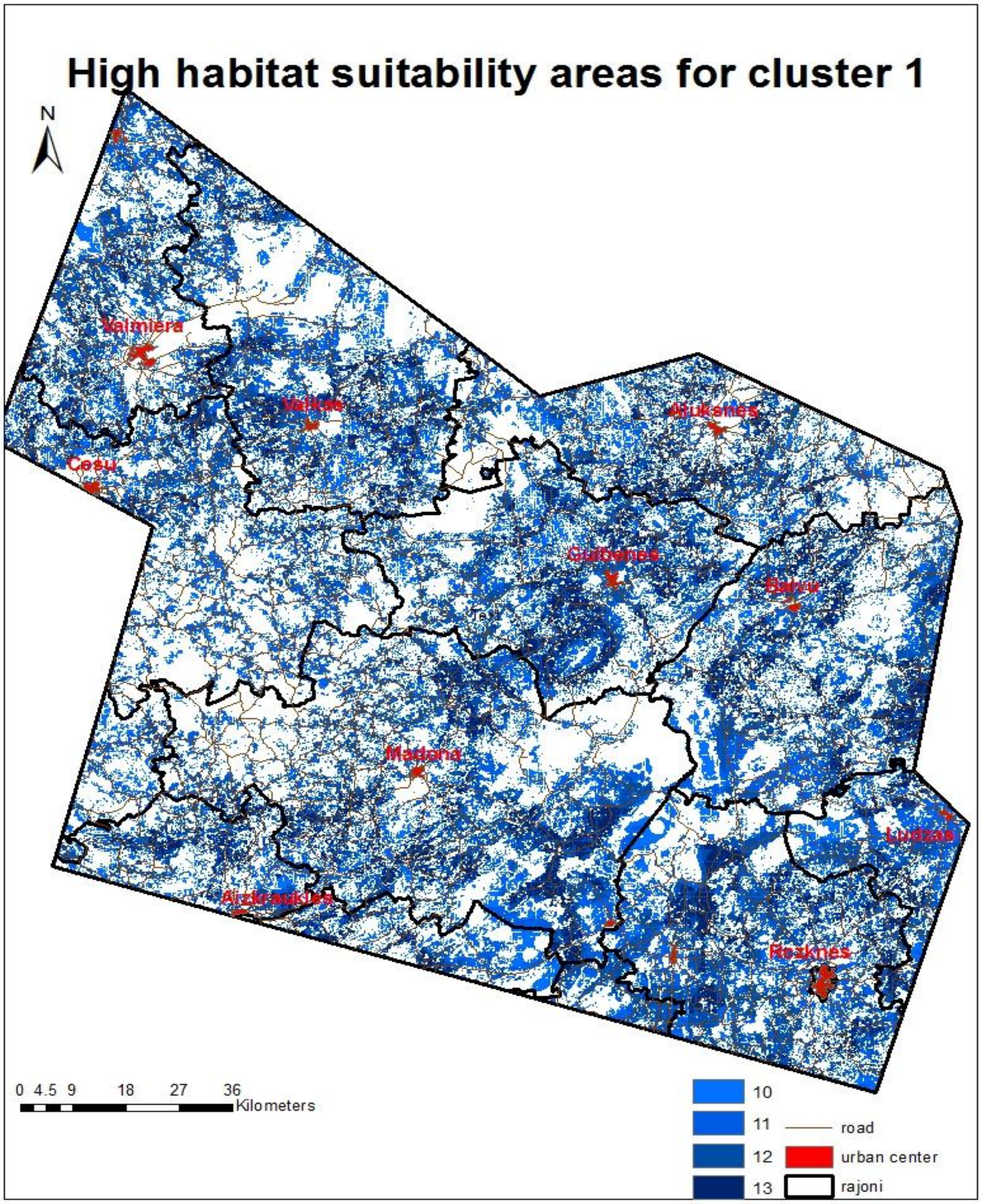

Fig. 3 Cluster 1 shows 10 out of 13 of all layers. These are areas where at least 10 out of 13 layers had suitable conditions for Giant Hogweed to grow.

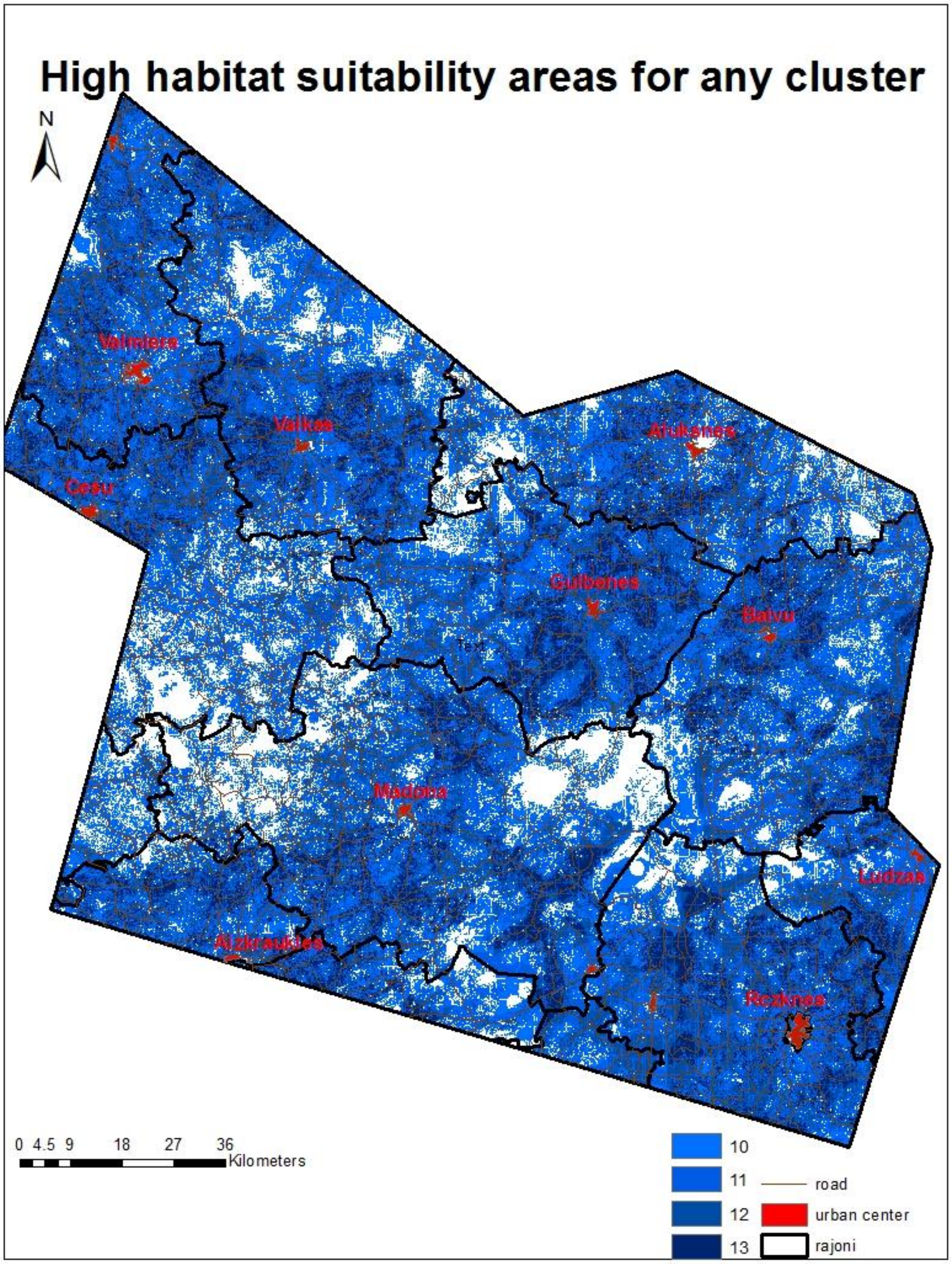

Fig. 4 This map shows all pixels whose environmental variables resemble any of the four clusters (based on having 10 or more variables fall within the 10^{th} – 90^{th} percentile ranges), and are therefore likely candidates for future Giant Hogweed colonization.

Discussion

The use of environmental, geographic and population data in predicting species distribution have been reported in other studies (Bethany et al. 2006; Mackey et al. 1988, 1989; Belbin 1993, 1995; Faith and Walker 1996a, b). This study examines the performance of cluster analysis in selecting a set of suitable environmental conditions where Giant Hogweed is present. The performance of the model was evaluated using an independent testing data set not used for training.

A cluster analysis was run on 377 cases, each responding to a set of environmental variables. A K-cluster analysis using Ward's method produced four clusters, for which the majority of the variables were significantly different. The final cluster centers are used to describe the clusters depending on how far the values are above or below the overall means (note all means are 0 since variables were standardized using Z-scores). These clusters represent the various environmental conditions where Giant Hogweed thrives in this landscape. The clusters were characterized by:

1) Cluster 1: Occurs in areas with moderate elevation with low slopes and where the predominant land cover type is agriculture near low population centers.
2) Cluster 2: Occurs in areas of low elevation and low slopes, near roads and urban centers, and in neighborhoods with high urban density and in neighborhoods where land cover change to both agriculture and urban is high.
3) Cluster three: Occurs in areas of low elevation and low slope, close to urban centers, in neighborhoods of high forest proportion and where where land cover change to forest is common, and near high population centers.
4) Cluster four: low elevation with steep slopes, near roads, near urban centers and close to water bodies and high population centers. The predominant neighborhood

land cover type in this cluster is agriculture and this cluster occurs in areas with little or no neighbourhood land cover change.

The distribution of clusters as shown on the cluster habitat suitability maps are representative of a set of environmental conditions that promote the occurrence of Giant Hogweed. From table 5, Cluster 1 has the highest number of cases (252) and contains the second largest area (48.5%) considered suitable for Giant Hogweed (table. 11). Cluster 3 has only 39 cases but has the highest largest area (53.22%) in terms of suitability. Cluster 2 and cluster 4 have the smallest areas considered suitable for Giant Hogweed occurrence.

Habitat suitability maps show areas that more likely to be suitable for Giant Hogweed. Figure 3, shows as example cluster 1 with low-high ranking from 0 – 13, with 0 being bad and 13 highly suitable. When only 13 is selected for all four clusters and tested using test data only 27% of the test data fall within the area considered suitable for Giant Hogweed. When 12 – 13 and 10-13 are selected and tested about 85% and 95% of test data fall within the areas considered suitable for Giant Hogweed respectively. Figure 5, show the 95% of test data while Figure 6, shows where 5% of test data fall within areas considered unsuitable for Giant Hogweed. The areas consider to be unsuitable are located far away from roads and could be due to bias in sampling Giant Hogweed locational points in the field.

Table 11 Percent suitable area in each cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Less suitable area sqkm (0-9)	9368.92	17455.30	8509.20	11470.05
Suitable area sqkm (10-13)	8824.67	738.30	9684.40	6723.55
Total area sqkm	18193.61	18193.61	18193.61	18193.61
Percent suitable areas (sqkm)	48.50 %	4.05 %	53.22 %	36.95 %

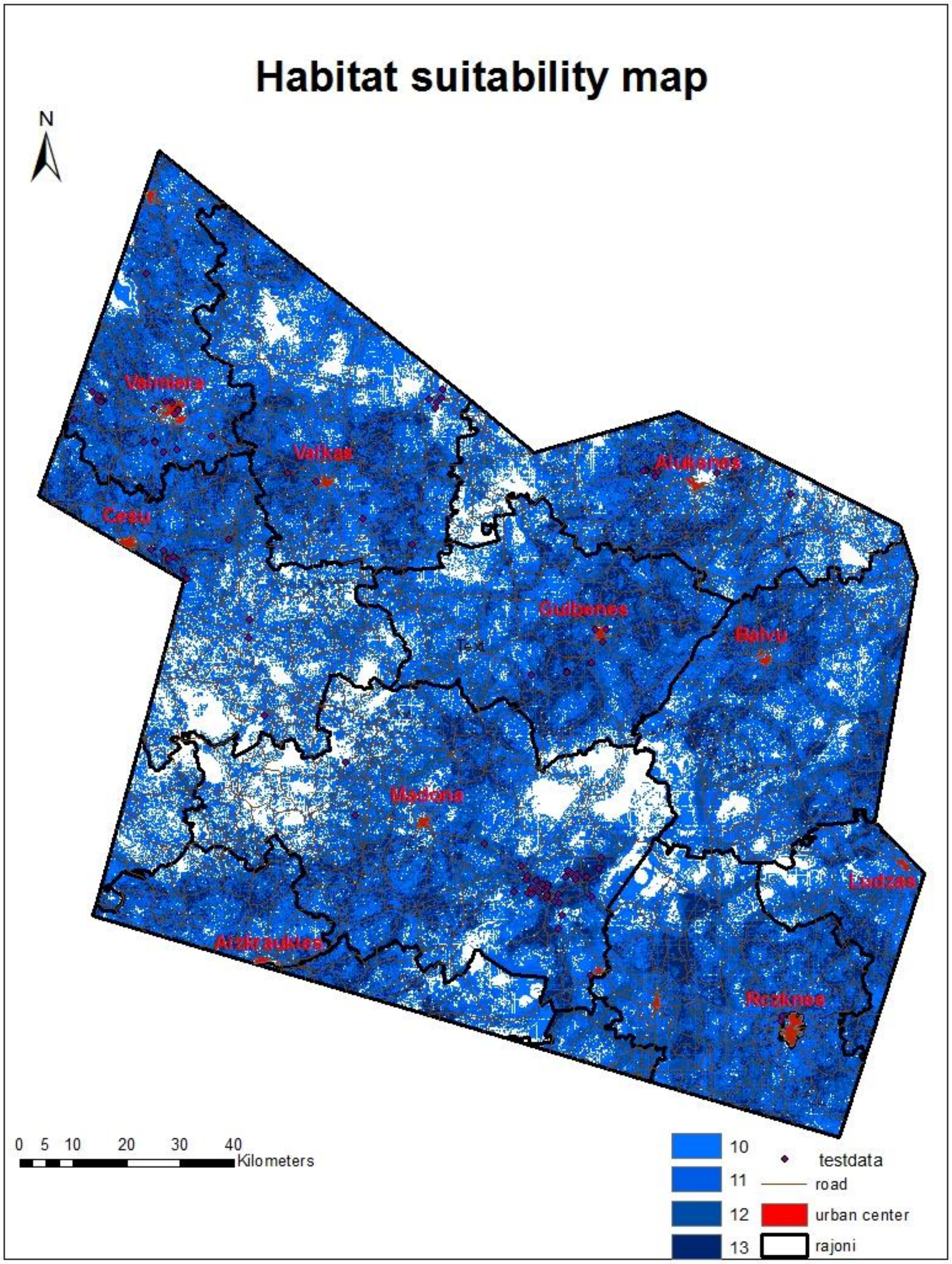

Fig. 5 95% of test data used to validate the model fall within the areas consider suitable for Giant Hogweed invasion, defined by suitability scores of 10 or greater.. 15.5 % of the total area is considered unsuitable.

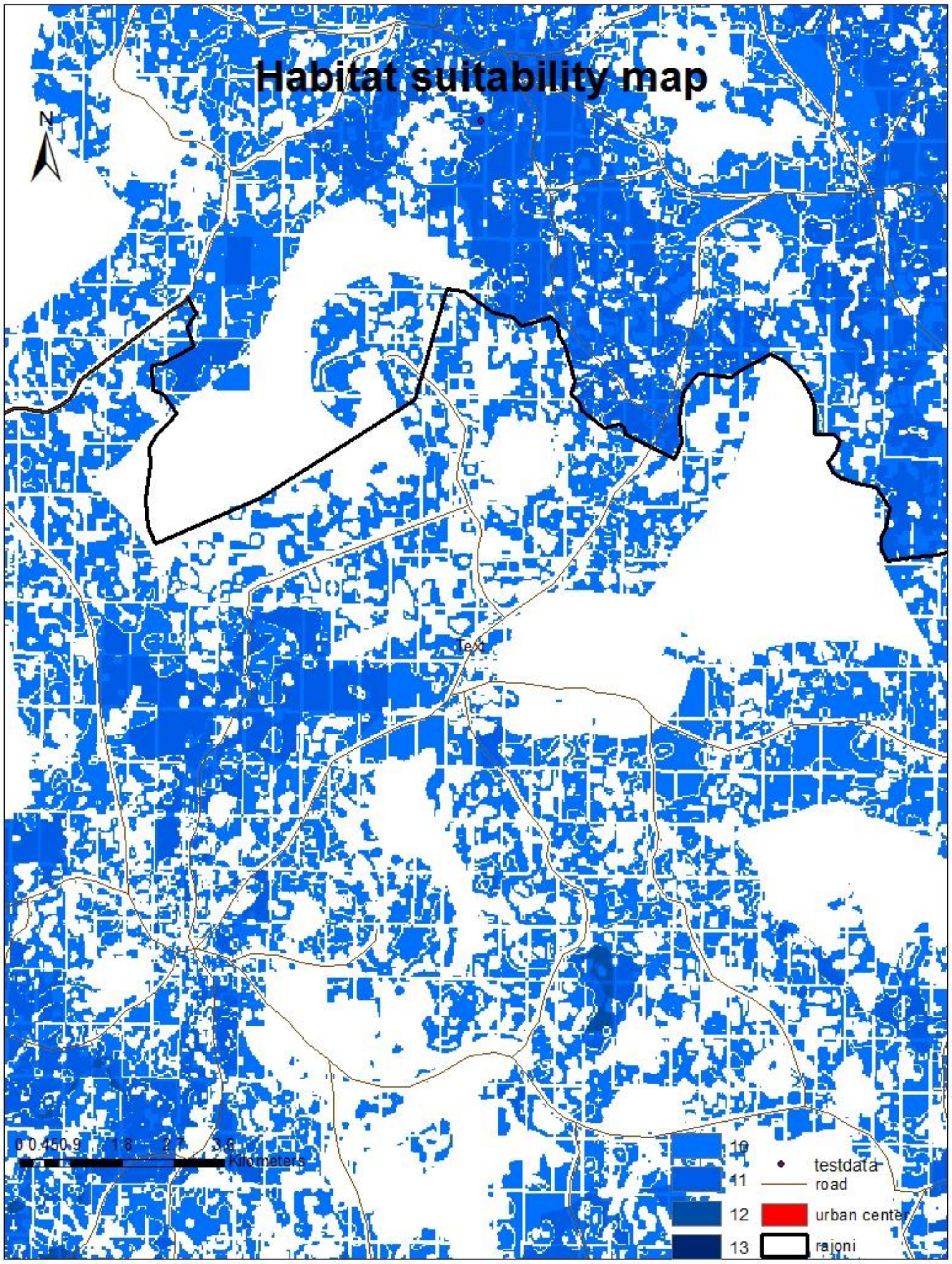

Fig. 6 Shows areas (white = less suitable) where just 5% of the test data fall (note the white, less suitable areas tend to be quite distant from roads)

Conclusions

Giant Hogweed *(H. sosnowskyi)* is included in the lists of the most aggressive alien species in all Baltic States (Gavrilova 2003). Therefore mapping and predicting the spatial distribution of this poisonous plant is very important for conservation planning and possible control/eradication efforts. The goal of this study was to evaluate the effectiveness of cluster analysis based on environmental factors as a new tool to select sites with favorable conditions for Giant Hogweed occurrence. Whereas most ecological niche models, e.g. MAXENT, can predict which environmental factors contribute the most or least in predicting species distribution, cluster analysis give different combinations of environmental factors that promote the spread of species. This analysis shows that for each factor (e.g., distance to roads) there is not necessarily one set of values that is the best predictor of Giant Hogweed locations – Giant Hogweed is found in different types of locations, each with its set of characteristic variables levels.

The results from cluster analysis indicate the following:

Cluster 1 occurs in mostly agricultural areas where the population density is low and little land-use change has occurred. As shown on the habitat suitability map, these areas are mostly rural areas where agricultural abandonment is common. During our field trip to Latvia in the summer of 2010, anecdotal evidence from discussions with farmers taught us that once farms are abandoned Giant Hogweed seeds regenerate and spread quickly.

Cluster 2 occurs mostly in urban areas close to roads and urban centers. This is probably near the larger cities where urban expansion is taking place and where disturbed land for new construction, including new roads, provides excellent habitat for Giant Hogweed.

Cluster 3: occurs near urban centers in neighborhoods with a high proportion of forest and where substantial land cover change to forest is occurring. These are probably less accessible areas near cities that are far from roads and along forest edges, likely where land has been abandoned.

Cluster 4 occurs mostly in urban areas near roads where the population density is high. These are most likely small farm areas and gardens in big cities, and roadsides. Cluster 4 is very similar to cluster 2.

This is the first known study that has utilized cluster analysis in this way to study habitat suitability. This study has shown that cluster analysis is not only useful to determine the multiple ecological/landscape niches a species can occupy (and has the ability to accommodate multiple optimal ranges of variable values), it also can be used to provide information on habitats suitable for Giant Hogweed spread. The combined cluster map shows the expected extent of potential invasion of Giant Hogweed in the study area. This map shows that the extent of potential invasion is generally higher in the lowlands than in the highlands and mostly in agricultural areas and along roadsides. This supports the findings of Pysek et al. (2007).

There are several factors that could contribute to the improvement of this approach. For instance other environmental factors such as temperature and rainfall and soil could be included in a future study, especially if a study is done on a more diverse, larger geographic extent. Z-scores were used in this study to run the K-cluster analysis, giving all the variables equal weight, but in reality some variables contribute more than others so future research may study ways to appropriately weight different environmental variables and to use more control data with all cluster types.

References

Almasi KN (2000) A non-native perennial invades a native forest. Biol. Invasions 2, 219–230.

Austin MP (2002a) Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. Ecological Modeling 157, 101-118

Austin MP (2002b) Case studies of the use of environmental gradients in vegetation and fauna modeling: theory and practice in Australia and New Zealand. In: Scott, J.M., Heglund, P.J., Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington pp. 64-73.

Belbin L (1993) Environmental representativeness: regional partitioning and reserve selection. Biological Conservation 66, 223-230.

Benezra C, Georges D, Yves S, Jean F (1985) Plant Contact Dermatitis. Toronto/Philadelphia.

Bethany AB, Mustard JF (2006) Characterizing the landscape dynamics of an invasive plant and risk of invation using remote sensing. Ecological Applications, Vol. 16, No. 3, pp. 1132-1147

Bohlen PJ, Scheu S, Hale CM, et al. (2004) Non-native invasive earthworms as agents of change in northern temperate forests. Front Ecol Environ 2, 427–35.

Bunkse EV (2000) Reality of Rural Landscape Symbolism in the Formation of a Post-Soviet, Postmodern Latvian. Identity. Norsk Geografisk Tidsskrift, Volume 53, pp. 121 – 138

Busby J (1991) BIOCLIM—a bioclimate analysis and prediction system. Plant Protection Quarterly 6, 8–9.

Clark JD, Dunn JE, Smith KG (1993) A multivariate model of female black bear habitat use for a geographic information system. Journal of Wildlife Management 57, 519–526.

D’Antonio CM, Vitousek PM (1992) Biological invasions by exotic grasses, the grass/fire cycle, and global change. Annu Rev Ecol Syst 23, 63–87.

Dukes JS, Mooney HA (1999) Does global change increase the success of biological invaders? Trends Ecol Evol 14, 135–39.

Eberhards G (1984) Reljefs (The relief). In: Latvijas enciklopedija (Encyclopedia of Latvia), Vol. 5, pp.34-41

Elith J., Graham C, NCEAS species distribution modeling group (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129-151.

Elith J., Leathwick JR (2007) Predicting species' distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Diversity and Distributions 13, 165-175.

Fairbanks DK, Benn GA (2000) Identifying re- gional landscapes for conservation planning: a case study from KwaZulu-Natal, South Africa. Landscape and Urban Planning 50, 237-257.

Faith DP, Walker PA (1996a) Environmental diver- sity: on the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. Biodiversity and Conservation 5, 399-415.

Gavrilova Ģ (2003) Report on investigation of *Heracleum* in Latvia in 2003. - Institute of Biology.

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993–1009.

Hirzel AH, Hausser J, Chessel D, Perrin N (2002) Ecological-niche factor analysis: How to compute habitat-suitability map without absence data. Ecology 83, 2027-2036.

Kaufman L, Rousseeuw PJ (2005) Finding groups in data. An introduction to cluster analysis. Wiley, Hoboken, NY

Kirkpatrick JB, Brown MJ (1994) A comparison of direct and environmental domain approaches to planning reservation of forest higher-plant communities and species in Tasmania. Conservation Biology 8, 217-224.

Kumar S, Stohlgren TJ, Chong G (2006) Spatial heterogeneity influences native and nonnative plant species richness. Ecology, 87, 3186-3199.

Lambrinos JG (2000) The variable invasive success of Cortaderia species in a complex landscape. Ecology, 83, 518-529.

Leathwick JR, Elith J, Hastie T (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modeling of species distributions. Ecological Modeling 199, 188-196.

Legendre P, Legendre L (1998) Numerical ecology. Second edition. Elsevier Science B.V., Amsterdam, The Netherlands

Lehman A, Overton JM, Leathwick JR (2002) GRASP: generalized regression analysis and spatial prediction. Ecological Modeling 157, 189-207.

Lele SR, Keim JL (2006) Weighted distributions and estimation of resource selection probability functions. Ecology 87, 3021–3028

Lodge DM, Williams S, MacIsaac HJ et al. (2006) Biological invasions: recommendations for US policy and management. Ecol Appl 16, 2035–54.

Mack RN, Simberloff D, Lonsdale WM et al. (2000) Biotic invasions: causes, epidemiology, global consequences, and control. Ecol Appl 10, 689–710.

Mackey BG, Nix HA, Hutchinson ME, Macmahon JP, Fleming PM (1988). Assessing representativeness of places for conservation reservation and heritage listing. Environmental Management 12, 502-514.

Manel S, Dias JM, Buckton ST, Ormerod SJ (1999) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. Journal of Applied Ecology 36,734–747.

Mladenoff DJ, Sickley TA, Wydeven AP (1999) Predicting gray wolf landscape recolonization: logistic regression models vs. new field data. Ecological Applications 9, 37–44.

Nielsen C, Ravn HP, Nentwig W, Wade M (2005) The Giant Hogweed Best Practice Manual. Guidelines for the management and control of an invasive weed in Europe. Forest & Landscape Denmark, Hoersholm, pp. 44

Normunds P (1993) Black Alder Swamps on Forested Peatlands in Latvia Folia Geobotanica & Phytotaxonomica, 28(3) 261-277 Published by: Springer

Obolevicˇa D (2001) Hogweed and its distribution in Latvia. http:// www.lva.gov.lv/daba/eng/biodiv/lauks_latvanis_e.htm

Osborne PE, Alonso JC, Bryant RG (2001) Modeling landscape-scale habitat-use using GIS and remote sensing: a case study with great bustards. J. Appl. Ecol., 38, 458-471

Parker IM, Simberloff D, Lonsdale WM et al. (1999) Impact: toward a framework for understanding the ecological effects of invaders. Biol Invasions 1: 3–19.

Pearson RG et al. (2007) Predicting species’ distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34: 102_117.

Pearson RG, Dawson TP, Berry PM, Harrison PA (2002) Species: A spatial evaluation climate impact on the envelop of species. Ecological modeling 154, 289-300.

Peeters ET, Gardeniers JP (1998) Logistic regression as a tool for defining habitat requirements of two common gammarids. Freshwater Biology 39, 605–615.

Pimentel D (2002) Introduction: non-native species in the world. In: Pimentel D (Ed). Biological invasions: economic and environmental costs of alien plant, animal, and microbe species. New York, NY: CRC Press.

Pyšek P, Cock MJ, Nentwig W, Ravn HP (2007). Ecology & management of giant hogweed (*Heracleum mantegazzianum*). - CAB International, Wallingford, UK.

Silveri A, Dunwiddie PW, Michaels HJ (2001) Logging and edaphic factors in the invasion of an Asian woody vine in a mesic North American forest. Biol. Invasions 3, 379–389.

Simberloff D (2000). Global climate change and introduced species in United States forests. Sci Tot Environ 262, 253–61.

Stockwell D (2006) Improving ecological niche models by data mining large environmental datasets for surrogate models. Ecological modeling 192, 188-196

Stockwell DR, Peters DP (1999) The GARP modeling system: Problems and solutions to automated spatial prediction. International Journal of Geographical Information Systems 13, 143-158

Stohlgren TJ, Barnett D, Kartez JT (2003) The rich get richer: patterns of plant invasions in the United States. Frontiers in Ecology and the Environment, 1, 11-14.

Store R, Kangas J (2001) Integrating spatial multi-criteria evaluation and expert knowledge for GIS-based habitat suitability modeling. Landscape and Urban Planning 55, 79-93.

Strayer DL, Caraco NF, Cole JJ et al. (1999) Transformation of freshwater ecosystems by bivalves: a case study of zebra mussels in the Hudson River. BioScience 49, 19–27.

Thompson LM, van Manen FT, Schlarbaum SE, DePoy M (2006) A spatial modeling approach to identify potential Butternut restoration sites in Mammoth Cave National Park. Restoration Ecology 14, 289–296

Van Manen FT, Young JA, Thatcher CA, Cass WB, Ulrey C (2005) Habitat models to assist plant protection efforts in Shenandoah National Park, Virginia, USA. Natural Areas Journal 25, 339–350.

Vitousek PM, D'Antonio CM, Loope LL, Westbrooks R (1996) Biological invasions as global environmental change. Am Sci 84, 468–78.

With KA (2002) The landscape ecology of invasive spread. Conservation Biology 16, 1192-1203.

With KA (2004) Assessing the risk of invasive spread in fragmented landscapes. Risk Analysis 24, 803-815.

Concluding remarks

This dissertation has presented research on LCC and the spread of invasive nonindigenous species. The study illustrates how human interactions with the environment can have adverse consequences such as the proliferation of invasive species. The main goal of this study was to monitor LCC and how LCC and geographic variables influence the spread of Giant Hogweed in Latvia. These three papers not only shed light on factors that promote LCC and the spread of Giant Hogweed but also elucidate the consequences of such changes and potential impacts both in the present and near future.

The first part of the study demonstrates how empirical relationships between LCC and demographic and geographic variables can be established using remote sensing and GIS. An understanding of these empirical relationships is essential for mapping, monitoring, and predicting future land use land cover change (Bethany et al. 2006). The results from this study revealed that landscape pattern in this region have undergone severe changes since Latvia became independent in 1991. During the Soviet era (prior to 1991), agriculture was a primary industry in Latvia employing much of the population. The single most widespread change was change to forest due to agricultural abandonment. Most of these changes to forest occurred in rural pagasti where depopulation is a significant trend (many people have migrated to bigger cities in search for jobs and better opportunities) and many farms have been abandoned. The results also indicate that most LCC occurs near roads and relatively near to Riga. Part 1 of this dissertation has demonstrated the linkage between land use land cover change and both socio-demographic factors and geographic variables.

Part 2 and Part 3 of this study demonstrate the use of ecological niche models in predicting species distribution. While logistic regression models use both presence and absence data, cluster analysis can be used with only presence data in predicting Giant Hogweed distribution. Both results revealed environmental factors that promote the spread of Giant Hogweed. Both studies indicated that distance to roads, distance to urban centers, low elevation areas, and land use land cover change are important factors influencing the spread of Giant Hogweed in northeastern Latvia. In addition to these factors, the cluster analysis indicated population density is also an important factor affecting Giant Hogweed occurrence even though stepwise logistic regression finds population density to be insignificant. Cluster analysis was used for the first time in this context and offers the added value of showing the set of combinations of factors that promote the spread of Giant Hogweed. That is, clusters 1, 2, 3 and 4 represent regions characterised by different combination of environmental factors (discussion section). Both models also show habitat suitability maps showing areas that are more likely to be occupied by Giant Hogweed. These maps are very important in terms of monitoring and managing Giant Hogweed spread, and can be very useful for conservation planning, land managers and policy makers as to where to allocate resources for control.

In this study significant correlation was found between LULCC data and both demographic and geographic variables, which imply demographic and geographic conditions, played an important role in regional landscape change. Comparison between the pagasti (municipality) and rajoni (county) correlations suggest that the rajoni scale is too broad to catch the appropriate processes with regards to some of the demographic and geographic variables. LULCC, demographic and geographic factors influence the spread of Giant Hogweed in Latvia. The created Giant Hogweed suitability maps showing predicted potential spread may be useful for managing and controlling Giant Hogweed in Latvia. This research may be applicable to other regions with high Giant Hogweed occurrence, such as the larger Baltic region and Russia, Central and Western Europe, Northern USA and Canada.

There are several factors that could contribute to the improvement of the analyses in this research. For instance, in this land use land cover change analysis (Part 1), finer scale socio-demographic and economic variables were not used because of their lack of availability at the city scale, but these fine-scale variables likely had significant influence on land use change. For the species distribution modeling, climatic variables were not incorporated into these studies because there is not much climate variability within this study site and Giant Hogweed is well-adapted to the small range of existing variability within the study site. Another limitation of these Giant Hogweed distribution modeling studies is the limited number (2) of satellite images used in the time series, and the time gap between them. For instance, it is likely that some areas were forests in 1992, then cut down and farmed during the late 1990's, and then abandoned (for instance, in 2004): such areas would be identified in this research cut forest, whereas in reality its most recent land use change was agriculture abandonment. While a higher frequency of satellite image data is useful to capture more of the actual land use changes, such high frequency land use change data (with satellite images from several time points) can easily become unwieldy for analysis.

Furthermore, while the population data was studied at 2 scales (pagasti and rajoni), this study was unable to relate population data to land use change and Giant Hogweed presence at an individual level. It would be useful to study how out-migration from individual houses affects LULCC and Giant Hogweed presence; however this would require a significant data acquisition project.