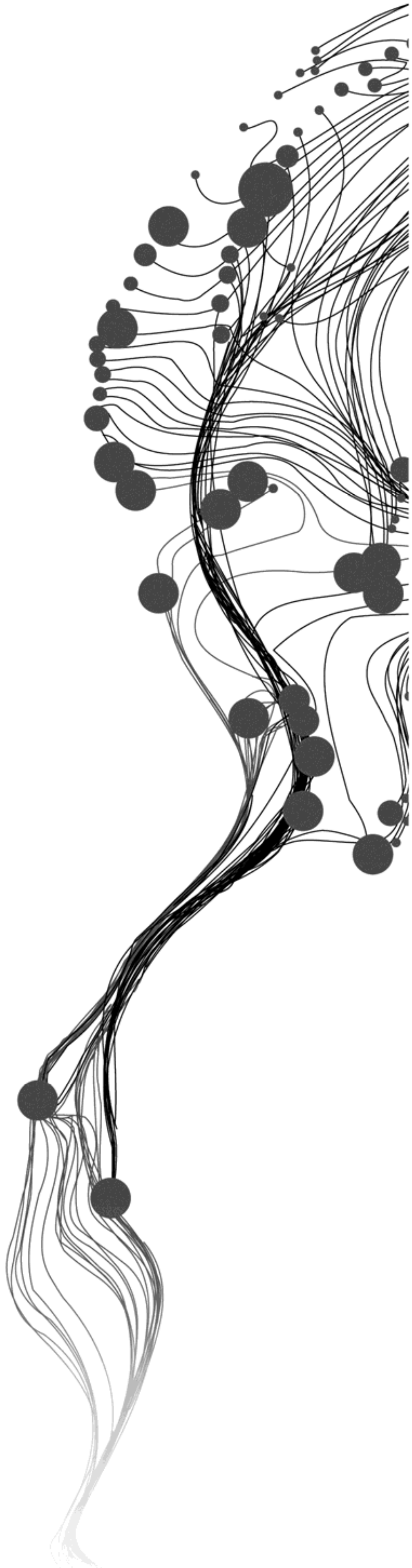
A COMPARISON OF LOGISTIC REGRESSION, GEOSTATISTICS AND MAXENT FOR DISTRIBUTION MODELING OF A FOREST ENDEMIC;

A PILOT STUDY ON LOBEL'S MAPLE AT MT. PIZZALTO, ITALY

ELHAM SUMARGA
March, 2011

SUPERVISORS:
Dr. Hein A.M.J. van Gils
Dr. David G. Rossiter



A COMPARISON OF LOGISTIC REGRESSION, GEOSTATISTICS AND MAXENT FOR DISTRIBUTION MODELING OF A FOREST ENDEMIC:

A PILOT STUDY ON LOBEL'S MAPLE AT MT. PIZZALTO, ITALY

ELHAM SUMARGA
Enschede, The Netherlands, March, 2011

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resource Management

SUPERVISORS:
Dr. Hein A.M.J. van Gils
Dr. David G. Rossiter

THESIS ASSESSMENT BOARD:
Dr. Bert Toxopeus (Chair)
Dr. J.F. Duivenvoorden (External Examiner, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam)

ABSTRACT

Acer cappadocicum subsp. *lobelii* (Lobel's maple) is a forest endemic which is native to southern Italy. Its population was observed at Mt. Pizzalto, Majella, Italy. A transect method was applied resulting in 31 presence points with 289 stems and 369 seedlings. The endemic *Acer* was distributed in a single patch of about 5.6 hectare. For modeling purpose, 87 absence points were also recorded across elevations and land cover types. Logistic regression, geostatistics and Maxent were used for distribution modeling of this endemic. Elevation, slope, aspect, annual incoming solar radiation, summer incoming solar radiation, land cover, NDVI, soil and geology were used as potential environmental predictors. A set of independent points consisting of 30% of the observed points was randomly selected for model validation. Geostatistics gave the best prediction with the perfect validation AUC (1). A stepwise Maxent provided a combination of elevation and aspect as the best and most parsimonious model with a validation AUC of 0.997. The highest probabilities to find this endemic are at the middle elevations (1450-1600 m) and southeastern expositions (80^{o}-190^{o}). The best logistic regression model consisted of slope, aspect, annual incoming solar radiation (ISR), soil and land cover as predictors with a validation AUC of 0.965. Of the predictors, only annual ISR is significant. The residuals of the three models were compared using ANOVA and there was no significant difference. The three models performed extremely well and gave very similar accuracies. Logistic regression and Maxent provided the potential distribution of this endemic while geostatistics predicted the actual one. For prediction over wider areas, Maxent appears preferable for its efficiency because requires presence-only data and fewer, easily obtainable environmental variables.

Keywords: Lobel's maple, forest endemic, logistic regression, Maxent, geostatistics, Mt. Pizzalto.

ACKNOWLEDGEMENTS

Finishing this thesis is a great achievement for me. I thank everybody who has supported me in accomplishing this hard job. Specially, I would like to send my gratitude to:

1. Dutch government, which trough Netherlands Education Support Office (NESO) Indonesia has awarded a StuNed scholarship for my study at ITC. This 15 month scholarship has brought me to “go beyond border”, and I am now ready to cross another border.
2. My first supervisor, Dr. Hein van Gils, who has guided me in all research phases. It was really not easy to design a good research, execute field data collection with unexpected problems and provide a scientific writing. His close “coaching”, both in the field and at ITC, supported me to pass through the difficulties.
3. My second supervisor, Dr. David Rossiter, for his valuable supervision on my thesis writing. His critical, thorough comments and suggestions were very constructive for improving my thesis. With his supervision, statistics was not frightening anymore.
4. Dr. Bert Toxopeus (Chair) and Dr. J.F. Duivenvoorden (External Examiner), for their willingness to examine my work.
5. NRM Programme Director, Dr. Michael Weir and all of ITC staff who have transferred their knowledge as well as offered great education services and atmosphere.
6. Dr. Giampero Csiachetti, for his assistance during fieldwork, especially for housing arrangement and discussion on the endemic *Acer*.
7. My fieldwork partners, Aida and Theresa, for a nice cooperation. We worked together, helped each other; and especially, they drove for me.
8. My national mate, Yanti, Andi *et al.* and all NRM 2009 students, who have given a very nice friendship, unforgettable moments and good cooperation.

Finally, my special thank goes to my wife, Mahmudah, who always gives me “sunshine” and never feel tired to pray for me. Also to my son, Faqih, and my daughter, Ikah, they are all my endless inspiration. To my parents, my brothers and my sisters, I thank them all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC = Akaike's Information Criterion
ANOVA = Analysis of Variance
AUC = Area under ROC Curve
BP = Before Present
BT = Transizione e Bacino
CRFA = Centro di Richerce Floristiche dell 'Apennino (Apennine Flora Research Center)
DBH = Diameter at Breast Height
DEM = Digital Elevation Model
GLM = Generalized Linear Model
GPS = Global Positioning System
ISR = Incoming Solar Radiation
JPEG = Joint Photographic Experts Group
Majella NP = Majella National Park
Maxent = Maximum Entrophy
NDVI = Normalized Difference Vegetation Index
QC1 = Continentale Pilo-Quaternaria 1
RMSE = Root Mean Square Error
ROC = Receiver Operating Characteristic
SDM = Species Distribution Modeling
VIF = Variance Inflation Factors

1. INTRODUCTION

1.1. Endemic Species

An endemic species is a species that naturally exists in a single geographical region and nowhere else (Primack, 1993). This term is an antonym of widespread or pandemic species which refers to species found in more than one geographical region (Watts, 1984). Further Watts (1984) categorized endemic species into broad endemics which could be in a continent scale and narrow endemics which have very small distribution ranges. With the focus on endemism in plants, Gentry (1986) used a sharper categorization of endemic species, *i.e.* wide endemic (distributional area up to 200,000 km^2), locally endemic (distributional area up to 75,000 km^2) and extremely endemic (range mostly 5 – 10 km^2). Dickore & Miehe (2002) used only one range limit for endemic plant, *i.e.* species with distributional area less than 100,000 km^2. However, without explicitly mentioning the area limits, Primack (1993) and Meffe & Caroll (1994) emphasized that the endemic term was most often applied to species with a very small geographical distribution. Many endemic species are isolated in specific region *e.g.* small islands, desert springs and mountain tops (Meffe & Caroll, 1994). In mountain regions, endemic species are mostly found at high elevation *e.g.* in Himalaya (Vetaas & Grytnes, 2002). In this mountain, 1957 out of 4928 species (40%) are endemic, most of them (around 600) distribute in the elevation range of 3800 – 4200 m.a.s.l. Tribsch (2010) also reported the high number of endemic plants in the Alps in which 249 out of 957 (26%) of high mountain plant species are endemic.

Geographic isolation of endemic species can be related to geological processes *e.g.* continental drift and mountain building (Meffe & Caroll, 1994), evolutionary process *e.g.* habitat specialization (Gentry, 1986) and climatic process *e.g.* sea level rising (Meffe & Caroll, 1994) and glaciations (Thompson, 2005). Continental drift and sea level rising have been identified as causes for high proportion of endemic species in islands. The well-known examples of the high endemism on an island are Hawaii (16,600 km^2) and Galapagos (7,900 km^2). Among 970 plant species found in Hawaii, 883 (91%) species are endemic. In Galapagos, 175 out of 701 plant species (25%) are endemic (Gentry, 1986). In relation to evolutionary process, Gentry (1986) identified that habitat specialization had resulted in the high local endemism of flora in some Amazonian families.

1.2. Endemic Species and Conservation Priority

Setting priorities is an important aspect in biodiversity conservation (Shi *et al.*, 2005) due to the limited conservation resources (Wilson, 2006) and uneven distribution of biodiversity (Mittermeier *et al.*, 1998). This prioritization needs to be established to achieve the effective conservation programs (Reid, 1998) and minimize biodiversity loss (Brooks *et al.*, 2006). Myers (2001) relates conservation priority to a question: "How can we save the most species at the least costs?" In another word, conservation priority also relates to the issue of how most of species richness can be saved by conserving only selected species of them (Bonn *et al.*, 2002).

Species endemism is one of the main criteria that are often considered in prioritizing conservation efforts (Peterson *et al.*, 1999). The other mostly used criteria are species richness and degree of threat (Bonn *et al.*, 2002; Myers *et al.*, 2000). These criteria are applied in determining regions with the high conservation priority both at global and national scales. At the global scales, Brooks *et al.* (2006) summarized nine schemes of global priorities for biodiversity conservation, *i.e.* biodiversity hotspots, crisis ecoregions, endemic bird areas, centers of plant diversity, megabiodiversity countries, global 200 ecoregions, high biodiversity wilderness areas, frontier forests and last of the wild. At national scale, the prioritization is mostly implemented in establishing conservation or protected areas (MacKinnon *et al.*, 1986), *e.g.* in India (Prasad *et al.*, 1998), Cameroon (Tchouto *et al.*, 2006) and Mexico (Leon-de la Luz & Breceda, 2006).

Recently, biodiversity hotspot is one of the most popular global priority schemes of biodiversity conservation. Myers (2001) defined biodiversity hotspot as a region with a very high concentration of endemic species and facing imminent threat of habitat destruction. Myers *et al.* (2000) identified 25 biodiversity hotspots around the world which totally contain 133,149 endemic vascular plants (44% of world's vascular plants). This describes the high importance of endemic species conservation.

Giving high conservation priority to endemic species means avoiding or at least retarding the species extinction rate. It is because endemic species is one of the species groups that are vulnerable to extinction (Primack, 1993). Due to the restricted geographical distribution, destruction of their habitat will more seriously threat their existence compared to widespread species. The extinction threat will be very high for very narrow endemic *e.g. Centaurea corymbosa* with a distributional range of only 3 km^2 in southern France (Freville *et al.*, 2004). Malcolm *et al.* (2005) analyzed the effects of global warming to extinction of endemic species from biodiversity hotspots and predicted the potential loss of around 56,000 endemic plants under doubled-CO^2 climates. Endemic species also tend to have low levels of natural competition (Primack, 1993), strengthening the need of taking care their existence in nature. Medail and Verlaque (1997) confirmed this issue by studying the ecological characteristics of endemic plants in southeast France and Corsica. They found that the distribution and frequency of endemics were correlated with three factors, *i.e.* specialized habitat, altitude and area with low interspecific competition.

1.3. Species Distribution Modeling

Providing information on spatial distribution of endemic species is a basic requirement for their conservation. However, intensive survey on those species is time consuming and expensive, demanding species distribution modeling (SDM) as an alternative (*e.g.* Hernandez, *et al*, 2008). SDM is a correlative model for predicting species distribution by explaining the observed pattern of species occurrence using environmental and/or geographic information (Elith & Graham, 2009). Besides for conservation biology, SDM is widely applied for studies on biogeography, ecology, and wildlife management (Araujo & Guisan, 2006). SDM may employ environmental variables (such as logistic regression and Maxent) or spatial autocorrelation (geostatistics).

Logistic regression is a type of regression when the response variable is binary and the explanatory variables can be either continuous or categorical (Hoshmer & Lemeshow, 2000). This model has been commonly applied in environmental and ecological studies such as in modeling deforestation (van Gils & Loza, 2006), potential beech expansion (van Gils *et al.*, 2008) and SDM (*e.g.* Manel, *et al.*, 1999 and Guisan *et al.*, 1999). In SDM, this model is applied by specifying species presence and absence data as response variable. Of generalized linear models (GLM), linear regression is the most frequently used for SDM (Rushton, *et al.*, 2004). This model uses the predictors to linearly model the logit of probability after which the probability can be generated using logit transformation (Takezawa, 2006).

Geostatistics is a general term for models of spatial structure that considers the spatial autocorrelation between values in the sampled locations; these models can then be used to predict the values of unsampled locations. The spatial autocorrelation analysis determines the correlation between values in certain separation distance (Isaak & Srivastava, 1989). This analysis is also called as variogram analysis from which a variogram model will be generated and then is used for prediction by kriging interpolation. Although geostatistics was initially developed for mining purposes (by D.G. Krige and G. Matheron) in 1960s (Webster & Oliver, 2007), its application have been spreading into various fields, including in ecology (Robertson, 1987). Geostatistical methods which are usually applied for SDM are indicator kriging (*e.g.* Guimaraes *et al.*, 2009) and regression kriging (*e.g.* Miller, 2005).

Maxent is one of the species distribution models that employs presence-only data as response variable (Baldwin, 2009). The input environmental variables for Maxent can be continuous or categorical (Philips & Dudik, 2008). Baldwin (2009) described at least three strengths of Maxent, *i.e.*:

- Requiring relatively small number of presence points to construct an accurate model.
- Less sensitive to spatial uncertainty of species records.
- Providing mapping features indicating whether a species is likely to be present or absent in a specific area.

In SDM, Maxent applies machine learning technique called maximum entropy (Hijmans & Graham, 2006). This model predicts the likelihood of species to occur by "finding the distribution of maximum entropy (*i.e.* closest to uniform) subject to the constraint that the expected value of each environmental variable under this estimated distribution matches its empirical average" (Philips *et al.*, 2004).

1.4. Problem Statement (Research Gaps)

In a biodiversity hotspot context, most regions of Italy are parts of the Mediterranean Basin hotspot. As a biodiversity hotspot, the Mediterranean Basin has high levels of species diversity and endemism. This hotspot is home to around 25000 plant species, 13000 of which are endemics (52%) (Myers, 2000). Those endemic plants are mainly concentrated on islands, peninsulas, rocky cliffs, and mountain peaks (Caley, 2008). Italy itself is the richest country in Europe on flora diversity, in which 6,759 species have been recorded in the Data Bank of Italian Vascular Flora (Wikipedia, 2010).

Majella National Park (NP) is one of the protected areas in Italy which is managed for biodiversity conservation. Situated in the region of Abruzzi, central eastern Italy, this park is the habitat of more than 2100 plant species which is equal to 65 % of Abruzzi flora, 37 % of Italian flora and 22 % of European flora. Among them, around 142 species (6.8%) are categorized as endemic species (Majella National Park, 2010). Habitat specialization might play main roles in forming endemism in this park as most of the endemics (67%) occur above tree line.

As part of conservation programs of endemic plants in Majella NP, some studies have been done to explore and model the spatial distribution of those species. For example, a database on endemic plant taxa and their presence points in this park has been being developed. Using this database, Nanyomo (2010) applied Maxent model to map the geographical distribution of five endemic plants in this park, *i.e. Campanula, Melampyrum italicum, Saxifraga italica, Bunium petraeum* and *Stipa dasyvaginata*. Chang (2010) also used this database to map the habitat suitability of 10 endemic plants in this park. Further, she created pseudo-absence points of two endemic species (*Achillea barrelieri* and *Adonis distorta*) and modeled their spatial distribution using indicator kriging.

By referring to the previous studies, it can be summarized that:

- Spatial distribution patterns of many endemic plants in Majella NP have not been modeled. It could be the matter of data sufficiency of those plants. Among 1560 records in database, only 194 points (12.44%) are at level-one spatial accuracy, *i.e.* recorded either by GPS or on the topographical map 1:25,000 with spatial uncertainty less than few tens meter.
- Studies on spatial modeling of endemic plants were based on presence-only data. Chang (2010) used absence points for her modeling, but the points were pseudo (not directly collected from the field).
- Endemic plant occurrences in forests were less observed. It is reasonable due to their distributional behaviour. However, some endemics are expected to occur in forests. A deeper study on their distribution in forests has not been conducted.
- Southern parts of Majella National Park, *e.g.* Mt. Pizzalto, were less studied in relation to the distributional patterns of endemic plants. This leads to the fewer records of those plants.
- Studies on endemics were focused on their distribution. An integrated study on their distribution and population structure has not been performed in Pizzalto.

By considering points mentioned above, the present study has as its objective modeling spatial distribution of forest endemics at Mt. Pizzalto, Majella NP. *Acer cappadocicum* Gled. subsp. *lobelii* (Ten.) Murray (endemic *Acer*) was selected as a modeling target because:

- It is a forest endemic and categorized as a rare species (Guarino & Napolitano, 2006). Although *Acer cappadocicum* widely spreads in Asia and Europe such as in Himalaya (*e.g.* Garkoti, 2008), China and Turkey (van Gelderen *et al.*, 2001), the subspecies *lobelii* is only found in southern Italy (Guarino & Napolitano, 2006; Di Pietro, 2009). Based on the records in CRFA (Centro di Richerce Floristiche dell 'Apennino) geodatabase, the distributional range of this endemic is about 2,100 km^2, categorized as narrow endemic (Watts, 1984) or locally endemic (Gentry, 1986).
- Only four occurrence points in Majella NP are recorded in database, and none of these has level-one spatial accuracy (Figure 1.1). Two records have level-two spatial accuracy which refers to toponym with spatial uncertainty less than 2 kilometers. The rest are in level-three spatial accuracy which refers to a spatial uncertainty less than 5 km.
- There was no data about its population density and population structure.

Pizzalto was chosen as study area because this endemic was expected to occur there. Additionally, Pizzalto can be called as a "microcosm" of Majella NP, since the variation of vegetation types along altitudinal zones in the entire park can be found in this mountain. This study was also directed to enrich the endemic data in this area. To test whether the use of presence-only data gives enough accuracy, absence points were also collected during fieldwork.

Logistic regression, geostatistics and Maxent models were selected to be applied. It was interesting to contrast the three models since they have different approaches in SDM. They also have distinct inputs in which logistic regression requires presence and absence data as well as environmental variables, geostatistics requires presence and absence data without environmental variables, while Maxent requires presence-only data and environmental variables. Elevation, slope, aspect, solar radiation, NDVI, land cover, soil and geology were involved as environmental variables. Those variables were suspected to have influences on the endemic *Acer* distribution. Plant distribution is usually limited by topographical conditions (elevation, slope, and aspect), solar energy availability (incoming solar radiation), vegetation community (land cover) and edaphic factors (soil and geology). NDVI was also involved to give a proxy of greenness variation to which the endemic *Acer* was expected to respond. Besides the accuracy, the successes of the three models were also evaluated in term of data requirements. As a pilot study, this research was projected to a broader application. Often, SDM is constrained by data availability, both on species occurrences and environmental variables. This study was also addressed to this issue.

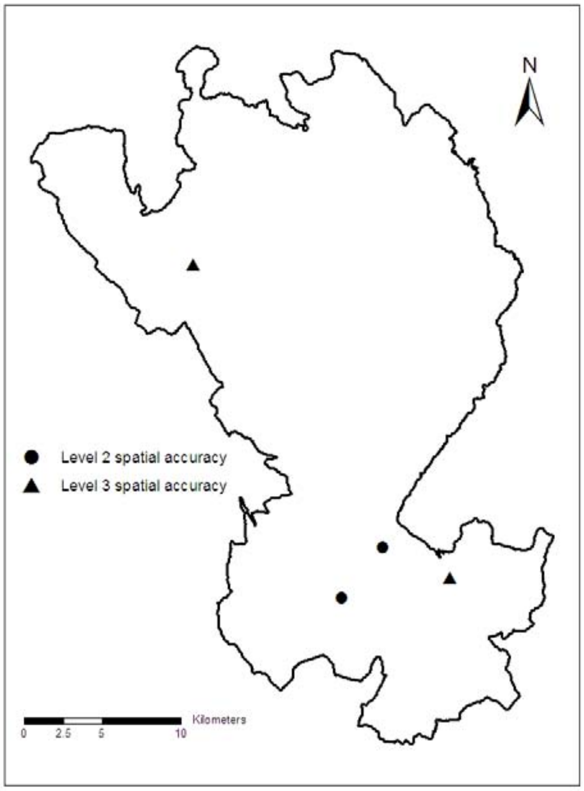


Figure 1.1. The endemic *Acer* records in Majella NP (level 2 spatial accuracy = spatial uncertainty up to two km, level 3 spatial accuracy = spatial uncertainty up to five km. Record in northern Majella is doubtful since this area is beyond the natural distribution of this endemic)

1.5. Research Objectives

The general objective of this research is to understand the distribution of the endemic *Acer* at Mt. Pizzalto and provide the best model for this endemic distribution. This objective can be broken down into some specific objectives as follow:

1. To identify the distribution and population structure of the endemic *Acer* at Mt. Pizzalto.
2. To model the distribution of the endemic *Acer* at Mt. Pizzalto using logistic regression, geostatistics and Maxent.
3. To compare the successes of the three models in term of model accuracy and parsimony.

1.6. Research Questions

In order to achieve the research objectives, some research questions were formulated as follow:

1. The observed endemic *Acer* distribution and population
    1.a. What is the distribution pattern of the endemic *Acer* at Mt. Pizzalto?

1.b. What are the environmental conditions of the endemic *Acer* distribution at Mt. Pizzalto?
1.c. What is the population structure of the endemic *Acer* at Mt. Pizzalto?

2. Distribution modeling of the endemic *Acer*
    2.a. Among the environmental variables (land cover types, NDVI, elevation, slope, aspect, solar radiation, soil types and geological types), what combination will best explain the presence probabilities of the endemic *Acer* at Mt. Pizzalto using logistic regression?
    2.b. Is there a spatial structure of this endemic occurrence at Mt. Pizzalto? If yes, how well the presence probability of this endemic in unsampled areas can be predicted by the spatial structure?
    2.c. How well can the presence probabilities of the endemic *Acer* at Mt. Pizzalto be predicted by using presence only data with Maxent model?
    2.d. Does the use of presence and absence data give a better accuracy than that using presence-only data in predicting the presence probabilities of the endemic *Acer* at Mt. Pizzalto?
    2.e. Which model is the most parsimonious for a broader application?

1.7. Research Hypotheses

The hypotheses tested in this study are:

1. Logistic regression analysis
    Ho : Slopes of all predictors (land cover, NDVI, elevation, slope, aspect, annual solar radiation, summer incoming solar radiation, soil types and geological types) are equal to zero (No predictor affects the presence probability of the endemic *Acer*).
    H1 : At least one predictor's slope is not equal to zero (at least one predictor affects the presence probability of the endemic *Acer*).

2. Geostatistics
    Ho : The endemic *Acer* individuals are randomly distributed throughout Mt. Pizzalto.
    H1 : The endemic *Acer* individuals occur preferentially in clusters.

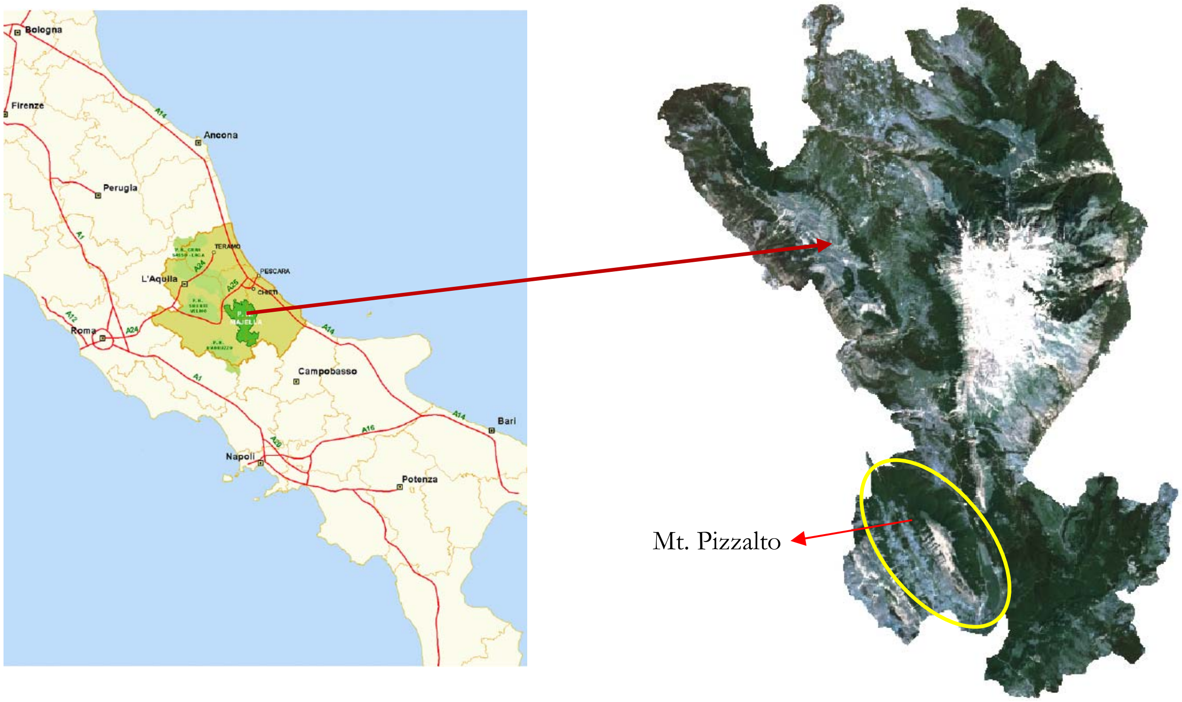
3. Maxent
    Ho : The best accuracy will only be obtained by involving all predictors (land cover, NDVI, elevation, slope, aspect, annual solar radiation, summer incoming solar radiation, soil types and geological types).
    H1 : The best accuracy can be obtained by involving fewer predictors.

4. Distribution model comparison (using ANOVA)
    Ho : There is no significant different of mean residual from different models (logistic regression, geostatistics and Maxent).
    H1 : At least there is a significant different of mean residual between two models.

2. MATERIALS AND METHODS

2.1. Study Area

This research took place at Mt. Pizzalto, Majella National Park, Italy. Located in Pescara, L'Aquila and Chieti provinces, this park covers area of 74,095 ha. Mt. Pizzalto is one of more than 60 mountains in this park, with the summit of 1939 meter above sea level. Up to 1875 meter, this mountain is dominated by beech forest, especially in the northern and eastern part. The study area was delineated based on the park boundary in northern part and the main road in the other parts.



Mt. Pizzalto

Figure 2.1. Study area

2.2. Materials

Maps and satellite images used in this research are listed in Table 2.1.

Table 2.1. Maps and satellite images used for the research

Data/Resolution	Source	Format	Data Type
Digital topographic map	ITC Majella database	JPEG	n.a.
Printed topographic map 1:50,000	Majella National Park	n.a.	n.a.
Land cover map 30 m	ITC Majella database	Raster	Categorical
Geological map 30 m	ITC Majella database	Raster	Categorical
Soil map 30 m	ITC Majella database	Raster	Categorical
DEM 30 m (2008)	ITC Majella database	Raster	Continuous

Slope map 30 m	DEM 30 m (2008)	Raster	Continuous
Aspect map 30 m	DEM 30 m (2008)	Raster	Continuous
Annual incoming solar radiation (ISR) map 30 m	DEM 30 m (2008)	Raster	Continuous
Summer ISR map 30 m	DEM 30 m (2008)	Raster	Continuous
Aster image 15 m (2008)	ITC Majella database	Raster	Continuous
NDVI map 15 m	Aster image (2008)	Raster	Continuous
Colour aerial photograph 0.5 m (2007)	ITC Majella database	Raster	Continuous

The maps and satellite images were used for two main purposes, *i.e.* for fieldwork and modeling. The topographic map and aerial photographs were used for fieldwork especially to identify the roads, tracks, cover types and elevation. Those were used both for sampling design and collecting data. Those were read in Arc Pad with a GPS to detect the position during data collection in the field. The maps used as inputs for modeling are land cover, soil, geology, elevation, slope, aspect, NDVI, annual ISR and summer ISR maps. The spatial resolution was defined at 30 m as it was the highest resolution available for most of the maps. Coarsening the resolution was not needed since all of presence and absence points were in level-one spatial accuracy. The following sections will describe the related information about the maps.

1. Land cover map

This map consists of 21 land cover classes. Since the endemic *Acer* was only found in beech forest, the map was reclassified into two classes, *i.e.* beech forest and non-beech (Figure 2.2) using the "reclassify" tool in spatial analyst tools in Arc Map (version 10) spatial analyst.

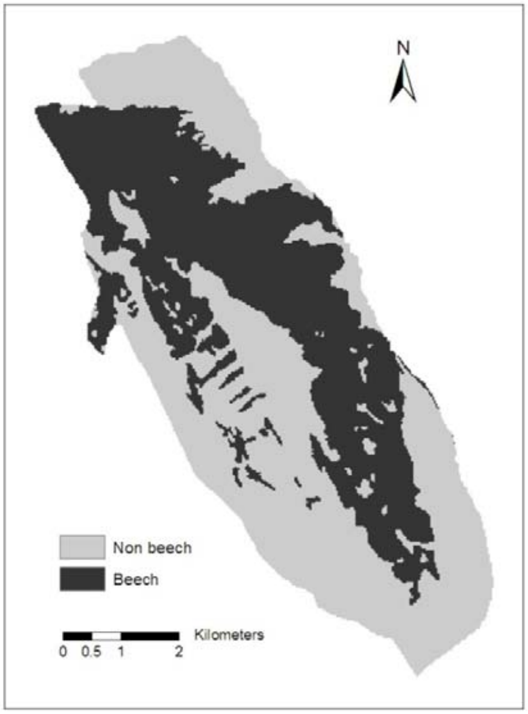


Figure 2.2. Land cover map of Pizzalto

2. Soil map

There are two kinds of soil map in the database. The first only contains the four main classes and the second one includes 18 sub soil classes. The map with four classes was used for this analysis. Because the endemic was only recorded in the "colluvial" class, the map was then reclassified into two classes, *i.e.* colluvial and others (Figure 2.3) using the "reclassify" tool of Arc Map.

3. Geological map

The endemic *Acer* was found in two of 12 geological classes, *i.e.* Transizione e Bacino (BT) and Continentale Pilo-Quaternaria 1 (QC1). Based on this, the map was reclassified into three classes, *i.e.* BT, QC1 and others (Figure 2.4) for the next analysis. The reclassification was performed using the "reclassify" tool of Arc Map.

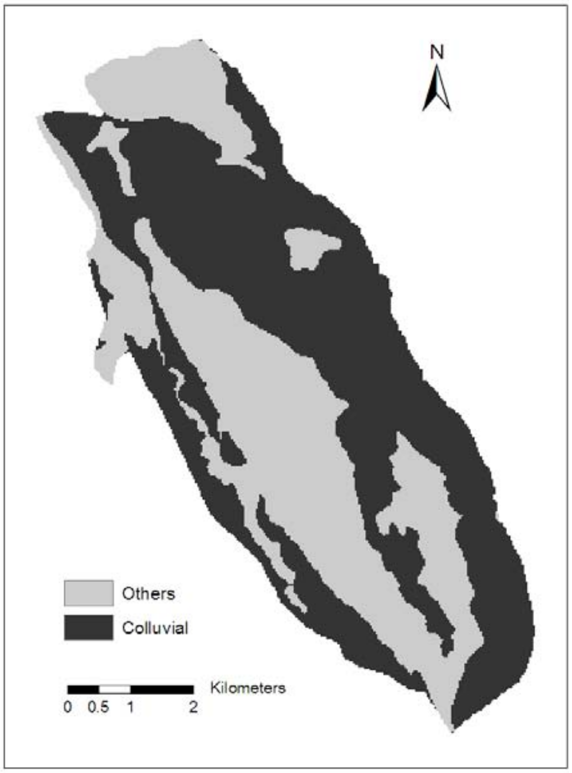


Figure 2.3. Soil map of Pizzalto

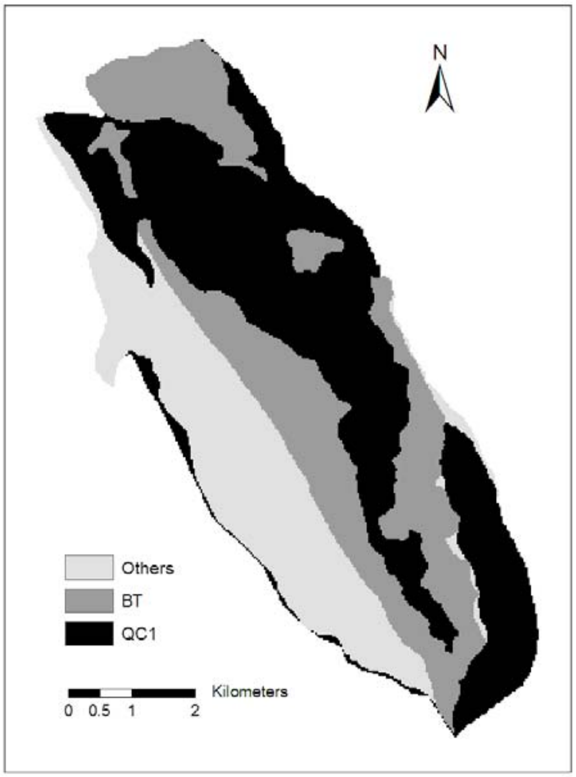


Figure 2.4. Geological map of Pizzalto

4. Elevation map

The elevation map (Figure 2.5) was directly created from the DEM without any modification.

5. Slope map

The slope map was generated from the DEM by using the "surface tool" in Arc Map spatial analyst. The map was expressed in degrees (Figure 2.6).

6. Aspect map

The aspect map was (Figure 2.7) also created from DEM by using the "surface tool" in Arc Map.

7. NDVI map

The NDVI map (Figure 2.8) was generated from the Aster image (2008). The "spectral enhancement" tool in "image interpreter" of ERDAS IMAGINE was used for this analysis.

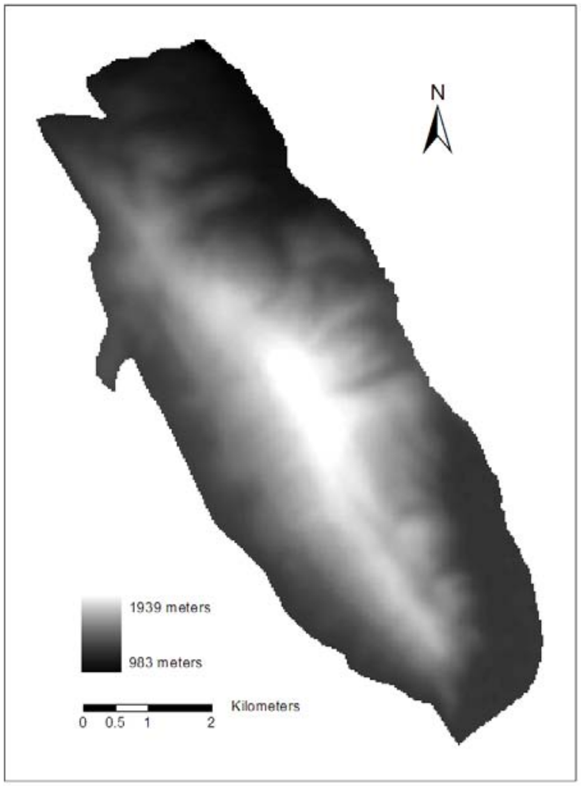


Figure 2.5. Elevation map of Pizzalto

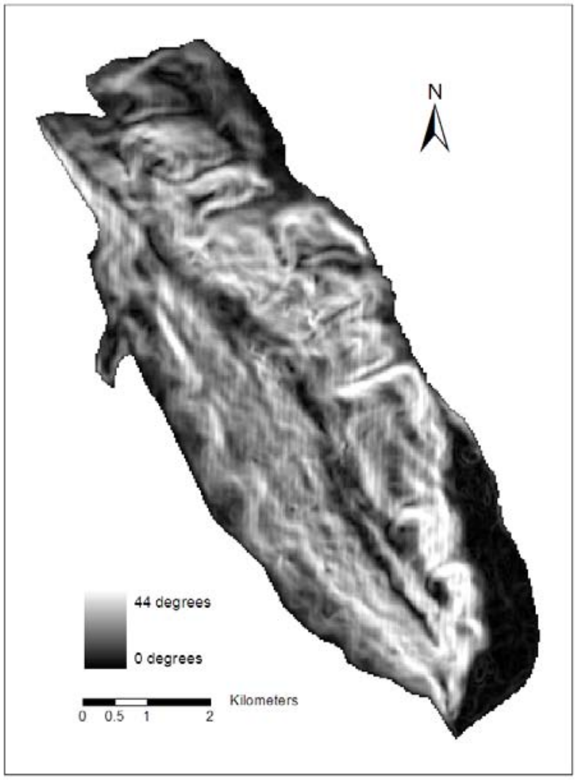


Figure 2.6. Slope map of Pizzalto

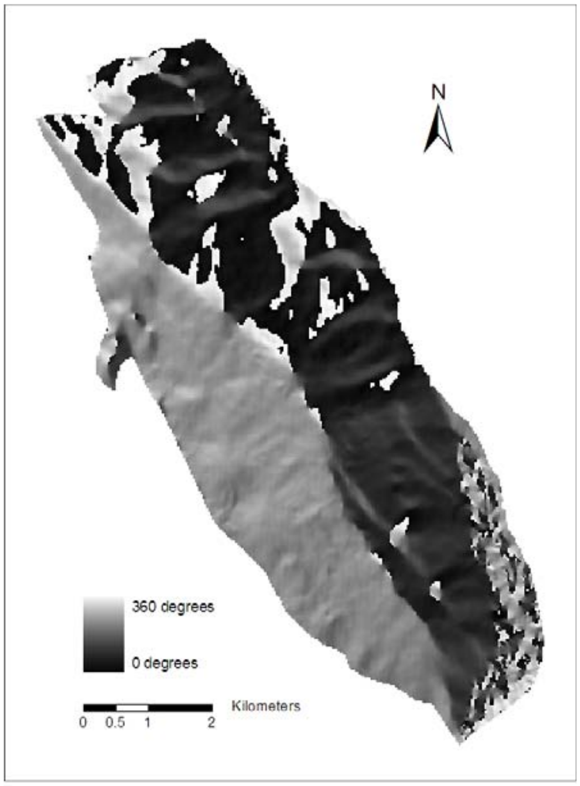


Figure 2.7. Aspect map of Pizzalto

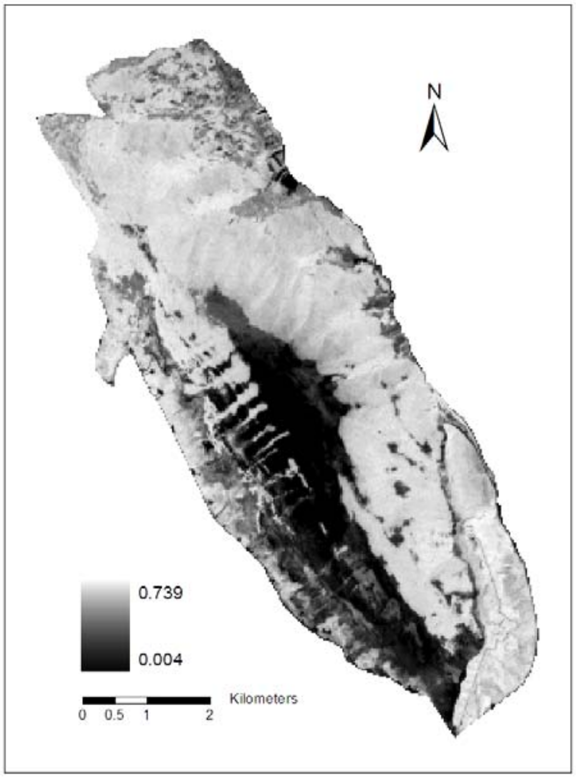


Figure 2.8. NDVI map of Pizzalto

8. Annual ISR map

The annual ISR map was generated from DEM by using the "solar radiation tool" in spatial analyst tools of Arc Map. Area solar radiation was selected for this analysis. The start day was set on the first of January 2010 and the end day was specified on 31th of December 2010. The map is shown in Figure 2.9.

9. Summer ISR map

This map was created in the same way as the annual ISR map except for the time which was set from the first of May 2010 to 31st of October 2010. The summer ISR map is given in Figure 2.10.

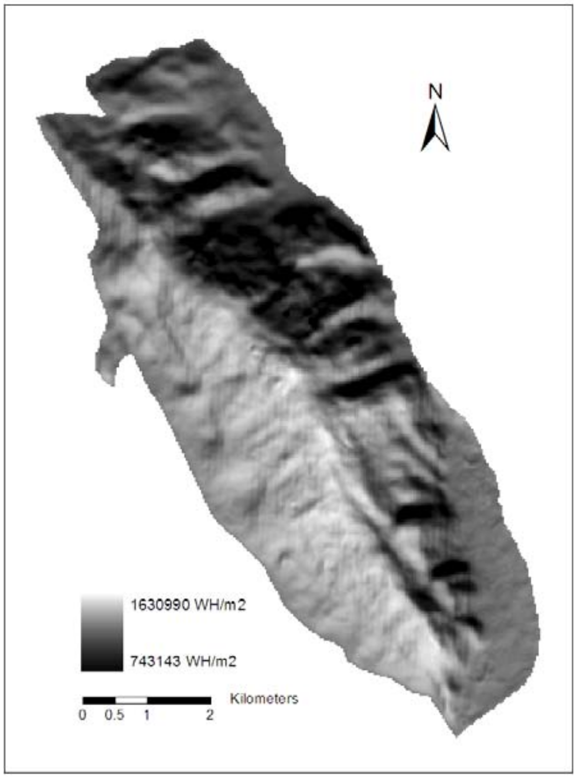


Figure 2.9. Annual ISR map of Pizzalto

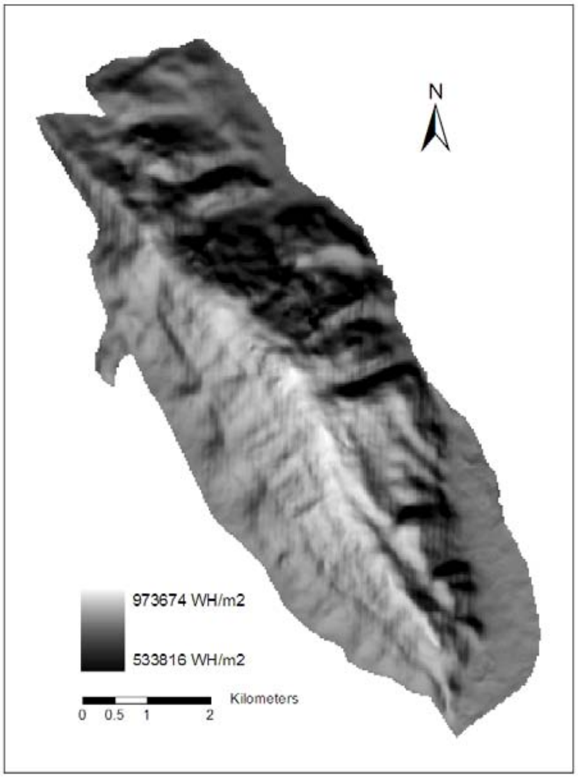


Figure 2.10. Summer ISR map of Pizzalto

2.3. Data Collection Method

A transect method was applied to collect the endemic *Acer* presence/absence data in Pizzalto. For optimizing the fieldwork time, twelve transects perpendicular to contour lines were completed (Figure 2.11). Considering the accessibility of study area, the transects were not exactly the straight lines with a systematic distance between transects.

The occurrences of the endemic *Acer* were observed along the transects. All individuals of this endemic found were recorded. Data recorded were XY coordinates, land cover type, number of stems, stem diameter and number of juveniles. The endemic *Acer* individuals which took place in the same pixel of 30 m were assigned as a single presence point. A different method was applied for absence point recording. Considering the high number of absence points along the transects, not all of absence points were recorded. The absence point data were recorded in points at 100 m elevation intervals. The difference in land cover types was also considered. More than one absence point were recorded in the same elevation interval if more than one land cover types were found in that interval. Only XY coordinates and land cover type were recorded in the absence points. The fieldwork was performed in a group of three.

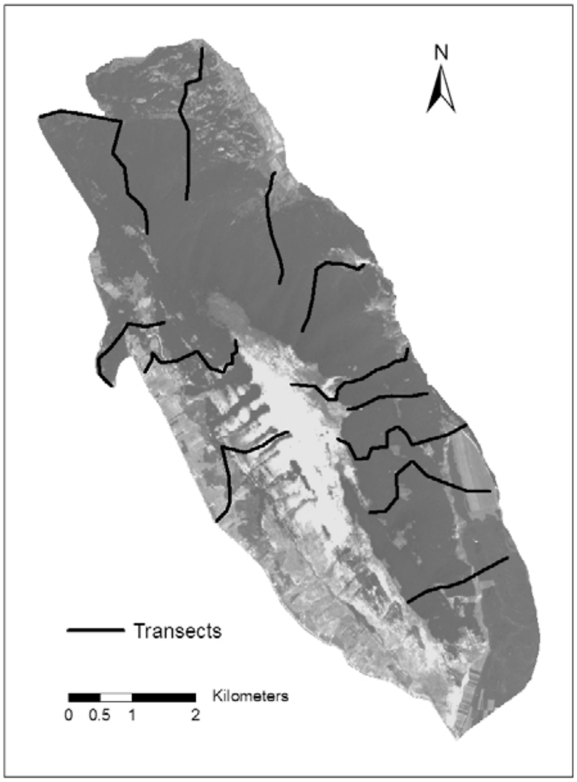


Figure 2.11. Transects for data collection

2.4. Data Analysis

Three models were applied for the distribution modeling of the endemic *Acer* in Pizzalto, *i.e.* logistic regression, geostatistics and Maxent. In order to make a fair comparison among the three models, 70% of data were randomly selected for modeling inputs and the rests (30%) were used for validation. So, the three models were validated by the same validation data. More information about the models is described in the following sections.

1. Logistic Regression

The GLM method of "R" with the binomial link family (Hastie & Pregibon, 1992) was used for the logistic regression analysis. The presence or absence of the endemic *Acer* was put as response variable while the explanatory variables were land cover types, soil types, geological types, elevation, slope, aspect, NDVI, annual ISR and summer ISR. A stepwise regression analysis was applied to get the best combination of predictors for the regression model. Before applying the stepwise procedure, a multi-collinearity analysis was done to analyze as well as to avoid correlation between predictors.

The general model for simple logistic regression is given in Formula 2.1.

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \tag{2.1}$$

The flowchart of logistic regression modeling is briefly described in Figure 2.12.

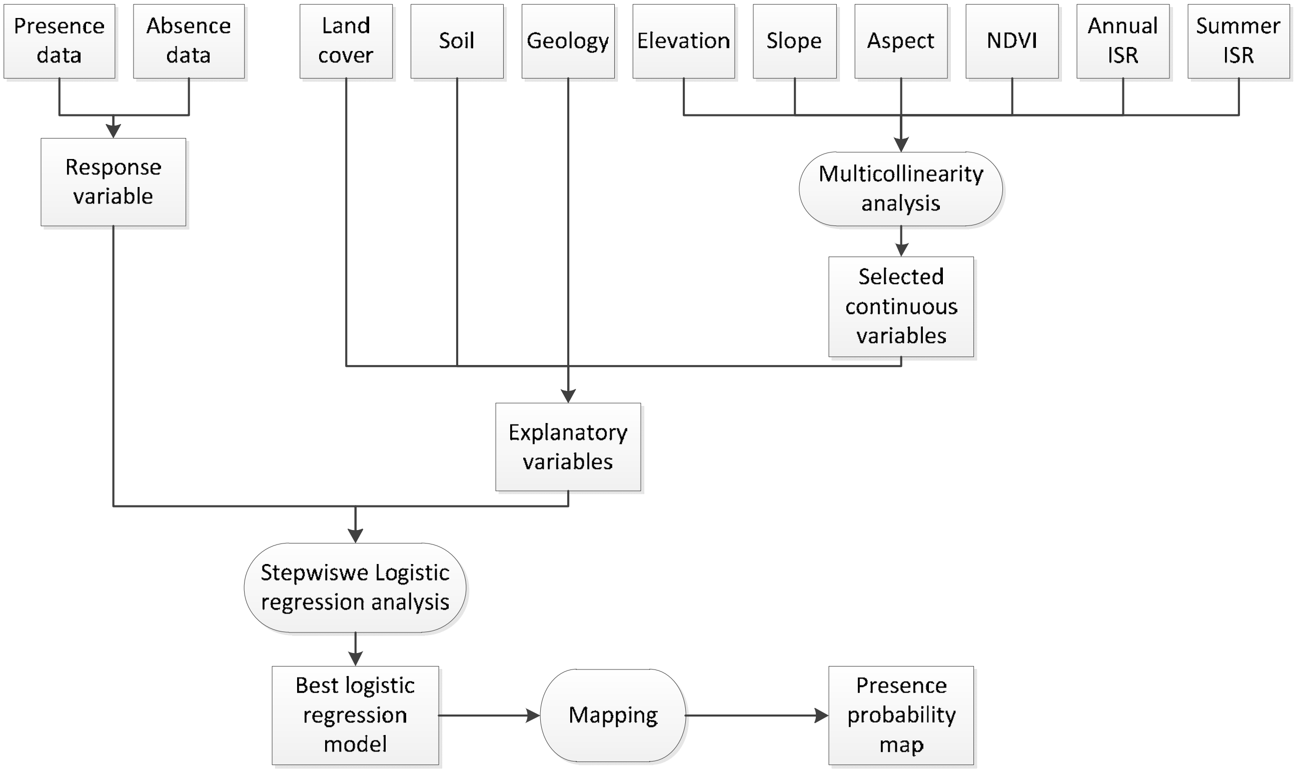


Figure 2.12. Flow diagram of logistic regression modeling

To measure the quality (the goodness of fit) of logistic regression model, a D^2 was used as an analogy of R^2 (Rossiter & Loza, 2010). The formula for calculating D^2 is given in Formula 2.2.

$$D^2 = 1 - (\text{residual deviance}/\text{null deviance}) \tag{2.2}$$

The residual deviance is the variability of the residuals while the null deviance is the variability of the dataset. So, a lower residual deviance leads to a higher D^2. For model comparison in stepwise logistic regression analysis, the Akaike's Information Criterion (AIC) was used as an analogue of adjusted R^2 (Rossiter & Loza, 2010). This value represents the adjusted residual deviance with the number of predictors. The better model is the model with smaller value of AIC because of its better fit to the data (Fox, 2002).

After resulting in the best regression model, the success (accuracy) of this model was described by measuring its sensitivity and specificity. Sensitivity is the proportion of true positive while specificity is the proportion of the true negative at a specific threshold for positive outcome (Rossiter & Loza, 2010). The sensitivity and specificity values in some different thresholds were then used to draw an ROC (Receiver Operating Characteristic) curve. ROC is a graph of sensitivity vs. (1- specificity) at different thresholds (Rossiter & Loza, 2010). Area under the ROC curve (AUC) describes the success of the model. The value of AUC is usually between 0.5 and 1. The AUC value of 0.5 means the accuracy of the model is equal to the random prediction while the AUC value of 1 indicates the perfect fit (Baldwin, 2009). The best model was also applied for mapping the presence probability of endemic *Acer* in Pizzalto. A raster calculator in spatial analyst tool of Arc Map was used for this mapping.

2. Geostatistics

Based on the type of input data, *i.e.* presence and absence data, an indicator kriging was selected as geostatistical technique. Indicator kriging produced the probability of the indicator at each location, *i.e.* from 0 to 1. The procedure of this technique is described in Figure 2.13.

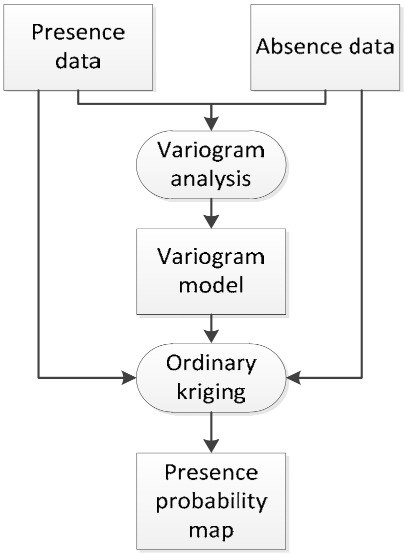


Figure 2.13. Flow chart of indicator kriging technique

Variogram analysis was performed to see the spatial structure of the presence and absence data. The gstat library of "R" (Pebesma, 2004) was used for this analysis. Variogram is a scatter plot describing the relationship between semivariance and distance. Ideally, the increase of separation between point-pairs will also be followed by the increase of semivariance until a certain separation called the range. After fitting a variogram model to the sample variogram, the parameters of the model were used for kriging interpolation. Ordinary kriging was selected as interpolation method. Ordinary kriging is a best linear unbiased predictor since it tries to have a mean error of zero and to minimize the variance error (Isaaks & Srivastava, 1989). An interpolation tool in spatial analysis tools of Arc Map was used for the interpolation to produce presence probability map of endemic *Acer* in Pizzalto.

3. Maximum Entropy (Maxent)

Maxent has two types of input, *i.e.* samples and environmental layers. In this analysis, the input for samples was the presence points of the endemic *Acer*. All explanatory variables (predictors) used in regression analysis were also used as the inputs for environmental layers. Maximum Entropy Species Distribution Modeling version 3.3.3 (Philips, 2010) was used for this modeling with the procedure shown in Figure 2.14.

There are three main results of Maxent, *i.e.* jackknife chart, ROC curve and probability map. Jackknife chart explains the contribution of each environmental layer to the final result. ROC curve describes the model's accuracy and the probability map shows the spatial distribution of predicted presence probability (Philips, 2010). A stepwise Maxent was performed with a backward method based on the result of jackknife test. A layer with a least contribution was removed and the model was rerun with the rest layers. Model with the highest AUC was considered as the best model. In case two or more models have the same AUC, model with fewer layers was selected.

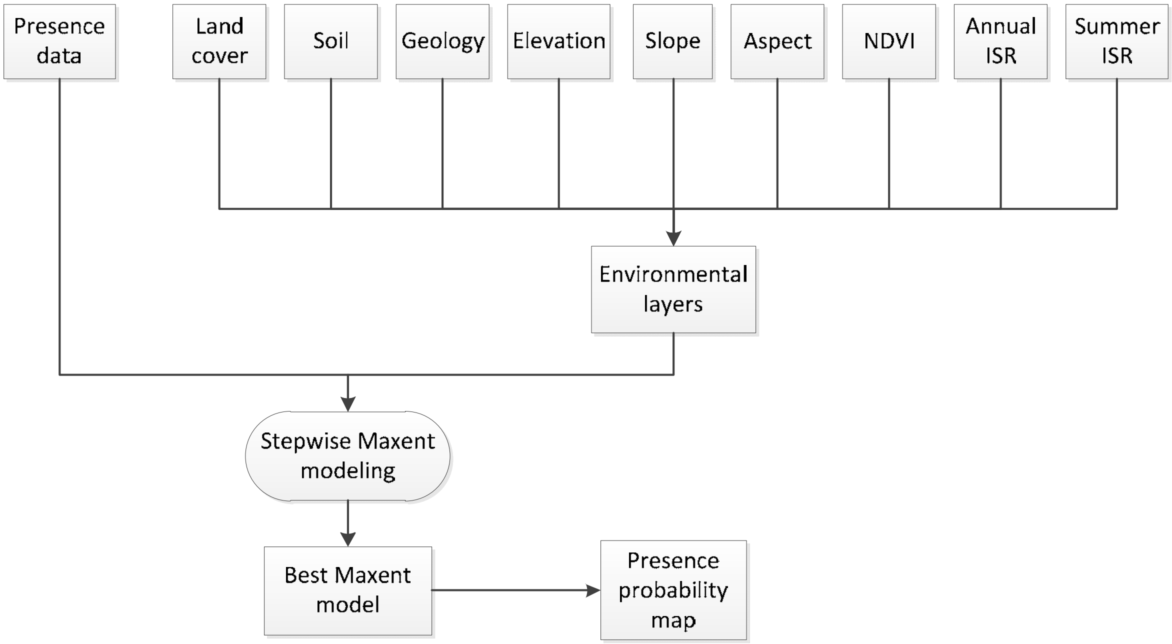


Figure 2.14. Flow chart of Maxent modeling

4. Model comparison

The best model of each distribution model (logistic regression, geostatistics and Maxent) were validated by the same independent validation data which consisted of 11 presence points and 26 absence points. The flow diagram of this analysis is given in Figure 2.15.

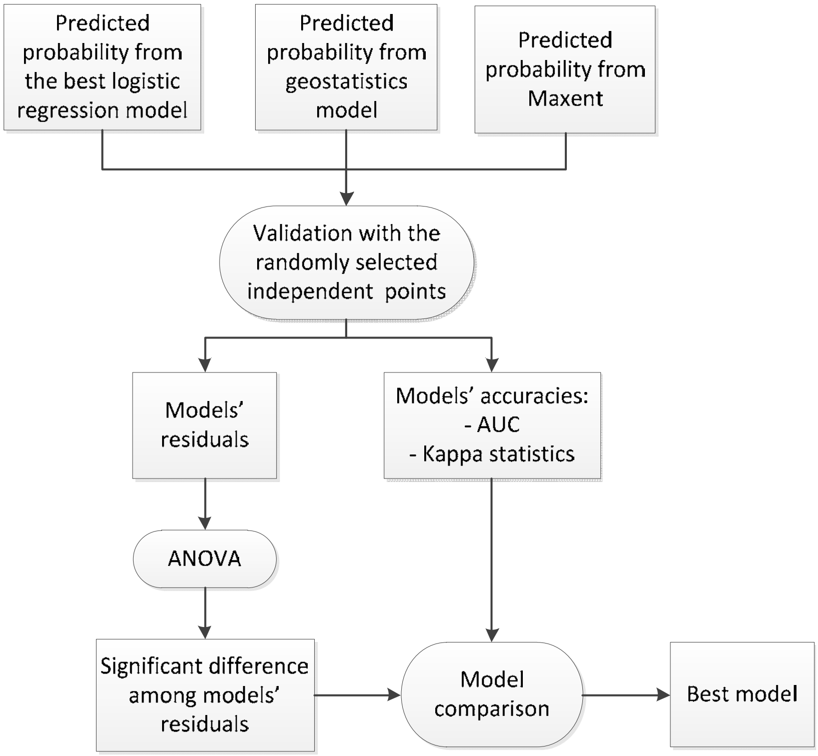


Figure 2.15. Flow chart of independent validation for model comparison

The successes of the three models were compared by analyzing their goodness in fitting their predicted presence probabilities and the independent validation data. The PresenceAbsence library of "R" (Freeman & Moisen, 2008) was used for this analysis. Two kinds of thresholds were used, *i.e.* single and multiple ones. For single threshold, a threshold at which the model gave a highest kappa was used as an optimal threshold. By using the threshold, the success of the model was measured by calculating the sensitivity, specificity, overall accuracy and kappa statistics. For multiple thresholds, an ROC curve was drawn and area under the curve (AUC) was calculated as a representation of model accuracy. The residuals of the three models were also analyzed using ANOVA to see whether there was a significant difference among them or not.

3. RESULTS

3.1. Observed Distribution and Population of the Endemic *Acer*

3.1.1. Distribution of Presence and Absence Points

Thirty-one (31) presence points and 87 absence points of the endemic *Acer* were collected during fieldwork (Figure 3.1).

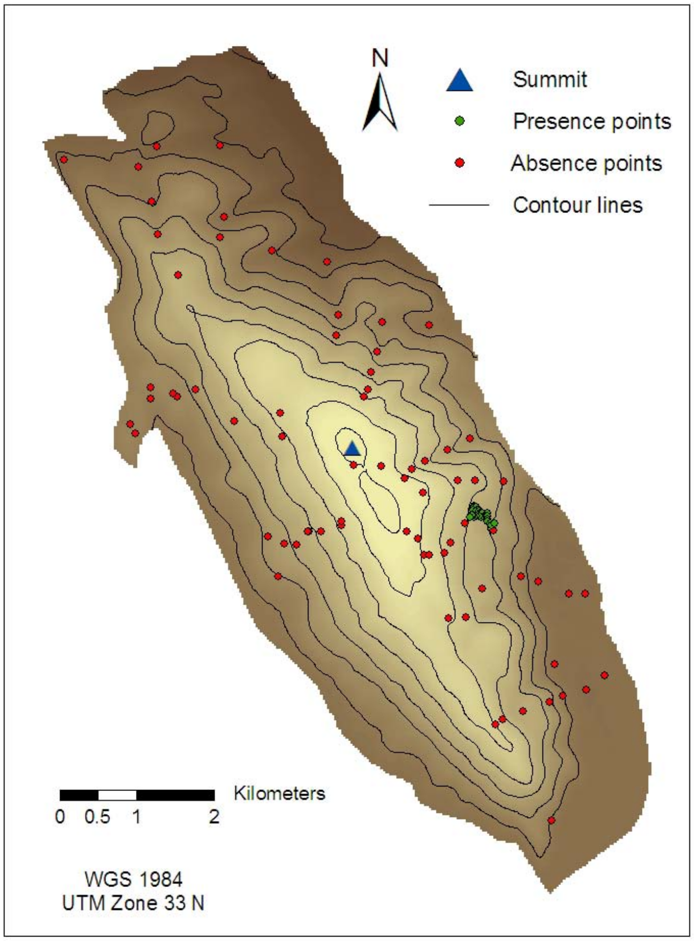


Figure 3.1. Distribution of presence and absence points of the endemic *Acer* at Mt. Pizzalto

All of the presence points were found in a single location (patch). By referring to the basic types of species distribution, *i.e.* uniform, clumped/aggregated and random distribution (Odum, 1971; Mauseth, 2009), the presence points of the endemic *Acer* at Mt. Pizzalto indicates a clumped distribution pattern. This is the answer of the research question 1.a. This pattern is shown by the aggregation of individuals living close together (Cotgreave & Forseth, 2002). The largest distance between presence points is 350 m. The area occupied by this endemic is about 5.6 ha.

3.1.2. The Endemic *Acer* and Environmental Variables

The following sections are addressed to the research question 1.b.

1. Elevation

The endemic *Acer* was found in the elevation between 1488 and 1590 m. It means the elevation range is 110 m only. This endemic distributes almost uniformly in the elevation range (Figure 3.2). This can be seen from the quite similar distance between minimum value, first quartile, median, third quartile and maximum value. Compared to the elevation range of Mt. Pizzalto (983 – 1939 m), the endemic *Acer* uses only 11.5 % of the range.

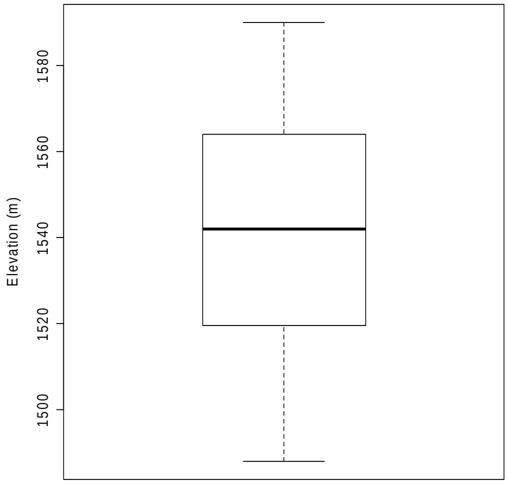


Figure 3.2. Boxplot of elevation range of endemic *Acer* at Mt. Pizzalto

2. Slope

The endemic *Acer* was found in the area with the slope between 8° and 26°. This can be categorized as gentle to medium slope. The slope distribution of the 31 presence points is described in Figure 3.3.

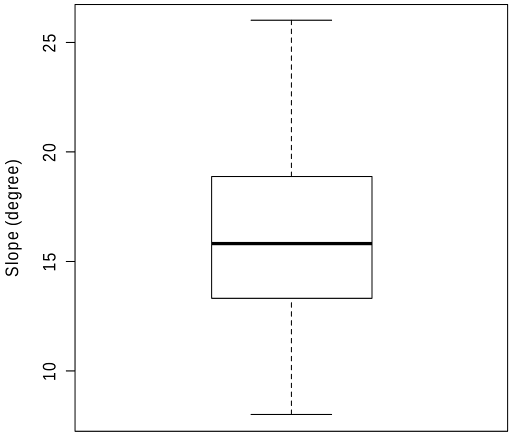


Figure 3.3. Boxplot of slope distribution of the endemic *Acer* presence points

There is a close distance between first quartile and third quartile. It means 50 % of the presence points distribute in the short slope range (13.32° – 18.85°).

3. Aspect

The presence points of the endemic *Acer* are located in the area with the aspect value from 84° to 150°. This range can be categorized into three directions, *i.e.* north to east, east to south and south to west. Most of the presence points (67.74%) of this endemic take place in east to south facing areas (Figure 3.4).

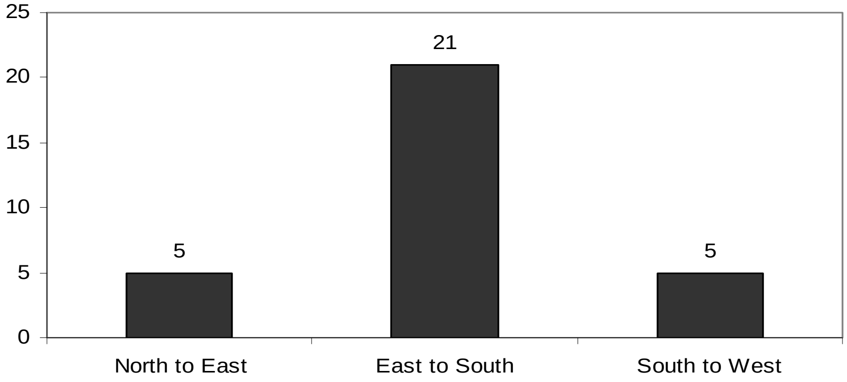


Figure 3.4. Number of presence points based on the aspect classes

4. Land cover

All of presence points of the endemic *Acer* were found in beech forest. As much as 77 % of this endemic grows under beech canopy and the rest are in the open. They are shade tolerant. Some of them even grow together with beech trees in the same point (Figure 3.5).

Figure 3.5. The endemic *Acer* and beech

5. Soil

All of presence points of the endemic *Acer* are located in the area with the same soil class, *i.e.* colluvial deposits mixed with debris and/or moraine. Among four main soil classes at Mt. Pizzalto, this class is the most dominant one that covers 56.5% area. Colluvium is material that was transported down and deposited by gravity and/or erosion (Spector, 2010), while moraine is material that was accumulated from glacial movement.

6. Geology

The endemic *Acer* was found in two of eight geological classes, *i.e.* Transizione e Bacino (BT) and Continentale Pilo-Quaternaria 1 (QC1). These two classes are the most dominant ones that cover 74.4% of the Mt. Pizzalto. The distribution of the presence points based on the geological classes is presented in Figure 3.6.

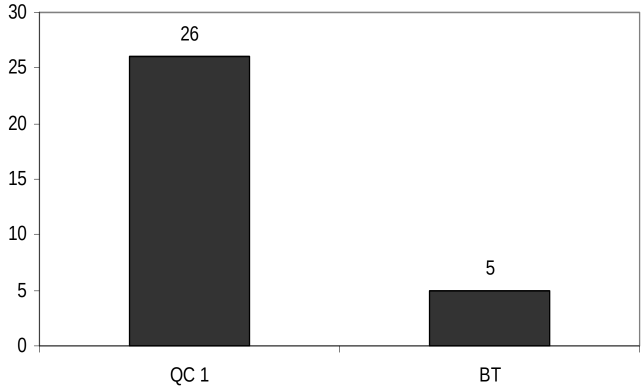


Figure 3.6. Number of presence points based on the geological types

3.1.3. Population Structure

The population structure of the endemic *Acer* in relation to its growth stages is listed in Table 3.1. In a horizontal bar chart, the population structure of the endemic *Acer* shapes a half pyramid (Figure 3.7). This is the answer of the research question 1.c. Some pictures of this endemic in different growth stages are presented in Figure 3.8.

Table 3.1. Population structure of the endemic *Acer* at Mt. Pizzalto

Growth stage	Height (m)	DBH (cm)	No individuals
Juvenile	< 1.5	< 10	369
Sapling	> 1.5	< 10	245
Pole	> 1.5	10 – 20	35
Tree	> 1.5	> 20	9

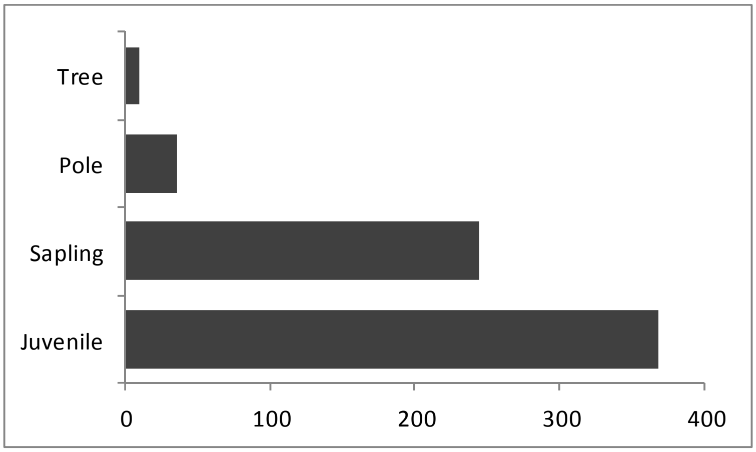


Figure 3.7. Shape of the endemic *Acer* population structure at Mt. Pizzalto

A B C

Figure 3.8. Pictures of the endemic *Acer* in different growth stages: juvenile (A), multistem with sapling and pole (B) and tree (C)

3.2. Distribution Modeling of The Endemic *Acer*

3.2.1. Logistic Regression Modeling

For multicollinearity analysis, the correlation coefficients between continuous input variables are summarized in Table 3.2.

Table 3.2. Coefficient of correlation between continuous input variables

	Elevation	Slope	Aspect	NDVI	Annual ISR	Summer ISR
Elevation	1	0.22	-0.17	-0.35	0.23	0.29
Slope		1	-0.25	0.06	-0.42	-0.55
Aspect			1	-0.21	0.38	0.34
NDVI				1	-0.19	-0.22
Annual ISR					1	0.97
Summer ISR						1

There are two coefficient of correlation with the value > 0.5, *i.e.* between summer ISR, annual ISR and slope. This indicates that multicollinearity may exist and will be a problem for the regression analysis. The variance inflation factors (VIF) of the variables are listed in Table 3.3.

Table 3.3. VIF value of continuous input variables

Variable	VIF
Slope	6.6
Annual Incoming Solar Radiation (ISR)	69.9
Summer ISR	94.6

Summer ISR has the highest VIF. This variable was predicted as the main factor causing the multicollinearity problem and was removed firstly from input variables. The VIF analysis after removing this variable showed that there were no more variables with the VIF value more than 10. It means the rest of input variables (elevation, slope, aspect, NDVI and annual ISR) are proper for multiple regression analysis.

The best five significant models resulting from stepwise logistic regression analysis are summarized in Table 3.4.

Table 3.4. Five best significant models resulting from stepwise logistic regression analysis

No	Explanatory variables	D^2	AIC
1	Elevation, slope, aspect, NDVI, annual ISR and soil.	0.72	39.7
2	Slope, aspect, NDVI, annual ISR and soil.	0.72	37.7
3	Slope, aspect, annual ISR and soil.	0.67	39.8
4	Slope, aspect, NDVI, annual ISR, soil and land cover	0.72	39.4
5	Slope, aspect, annual ISR, soil and land cover	0.72	37.5

Model-5 was selected as the best model because of its lowest AIC. The lowest AIC means the lowest adjusted residual deviance with the number of predictors. The D^2 of this model is the best. The D^2 of 0.72 describes that the model is capable of explaining 72% variation of presence probability of the endemic. The coefficients of the model's variables (Table 3.5) were then used to build a logistic regression formula (Formula 3.1)

Table 3.5. Coefficients and significances of predictors of the best logistic regression model

Predictors	Coefficients	P values	Significances at $\alpha = 0.5$
Intercept	- 56.36	0.002***	Very significant
Slope	0.221	0.10	Not significant
Aspect	- 0.042	0.12	Not significant
Annual ISR	0.000041	0.001***	Very Significant
Soil (non-colluvial)	- 20.89	0.99	Not significant
Land cover (non-beech)	- 19.94	0.99	Not significant

$$Y = \frac{1}{1+exp^{-(-56.36+(0.221*slope)-(0.042*aspect)+(0.000041*annualISR)-(20.89*soil.others)-(19.94*landcover.others)}} \qquad (3.1)$$

The coefficients and the significances of the predictors (Table 3.5) indicate the influence of the predictors to the presence probability of this endemic. By using the logit of presence probability as response variable, Formula 3.1 can also be expressed in a linear relationship (Formula 3.2).

$$Logit\ (P) = (0.221*slope) - (0.042*aspect) + (0.000041*annualISR) - (20.89*soil.others) - (19.94*landcover.others) - 56.36 \qquad (3.2)$$

where P = presence probability of the endemic *Acer*

The intercept of -56.36 explains that the logit of presence probability in beech forest with the colluvial soil type, slope of 0, aspect of 0 and annual ISR of 0 is -56.36. This is equal to presence probability of 3.34 e-25 and can be considered as zero probability. It is logical as there is no place in Pizzalto with the solar radiation of 0. The influence of each predictor is described in the following sections.

a. Slope

Slope has a positive relationship to the endemic presence probability. The higher the slope is, the higher the presence probability of the endemic *Acer* will be. One degree increase of slope will be responded by 0.221 increase of the logit of presence probability. However, this predictor is not significant at $\alpha = 0.5$. This means that the gradient of change in logit of presence probability in responding the slope change is statistically equal to zero. In other words, the slope variation doesn't significantly affect the variation of the logit of presence probability.

b. Aspect

Aspect has a negative relationship to the endemic presence probability. The logit presence probability will decrease 0.042 in every one degree increase of aspect. This predictor is also not significant at $\alpha = 0.5$, indicating there is no significant effect of aspect to the logit of presence probability.

c. Annual ISR

Annual ISR is the only predictor that is very significant at $\alpha = 0.5$. This means the change in annual ISR will significantly change the logit of presence probability. The relationship is positive in which one watt.hour/m^2 increase of annual ISR will be followed by 0.000041 increase of logit of presence probability. It seems the 0.000041 increase is very low, but actually the variation of annual ISR data is in 100 kilowatt.hour/m^2. So, we can also say that in every 100 kilowatt.hour/m^2 increase of annual ISR, the logit of presence probability will increase by 41. It is reasonable that this is considered as a significant change.

d. Soil

Because soil is categorical, logistic regression took one class as a reference with the coefficient of 0. Colluvial class was taken as the reference. In non colluvial soil type, the logit of presence probability will decrease 20.89. Based on the P value, the 20.89 difference of the logit between colluvial and non-colluvial type is not significant.

e. Land cover

Land cover is also categorical and beech forest class was taken as a reference. The logit of presence probability in non-beech forest is lower than that in beech forest with the different value of 19.94. However, the difference was not significant.

In summary, although only one predictor is significant, the combination of the five predictors gave the best accuracy in predicting the presence probability of the endemic *Acer* in Pizzalto. This is the answer of the research question 2.a. A combination of slope, aspect, annual ISR, soil and land cover best explained the presence probability of this endemic. The first research hypothesis is also answered. The null hypothesis, *i.e.* no predictor is significant, is rejected.

The success of this model can be described by measuring its sensitivity and specificity. Sensitivity shows the model's ability to predict the presence of the endemic while sensitivity explains the ability to predict the absence one. By using a threshold of 0.5, the model gives the sensitivity of 0.95 and specificity of 0.97 (Figure 3.9). It means that the model has good ability in predicting both the presence and absence of the endemic *Acer* at Mt. Pizzalto.

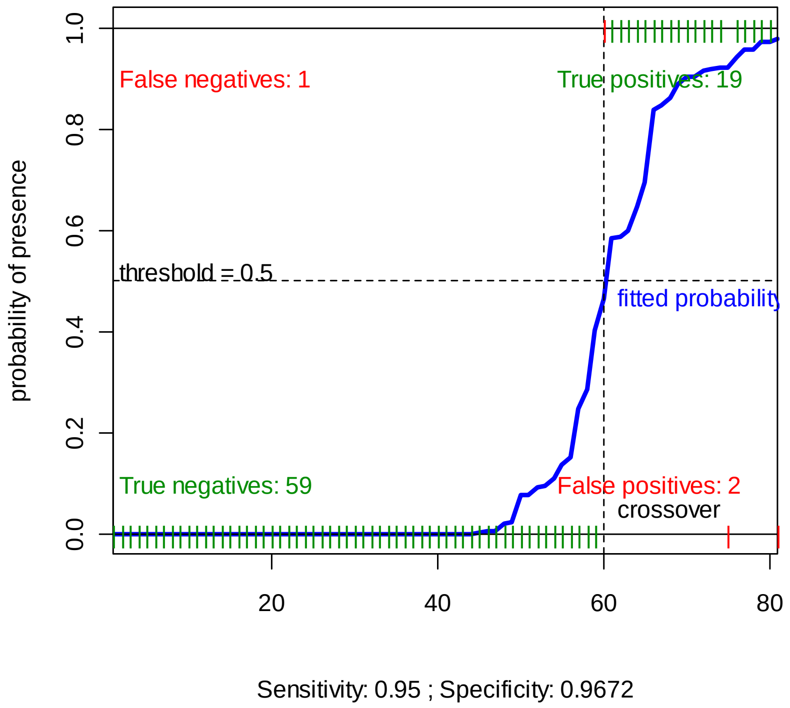


Figure 3.9. Sensitivity and specificity of model-5 with slope, aspect, annual ISR, soil and land cover as predictors

The success of the model was also evaluated by calculating the area under the ROC curve (AUC) with result shown in Figure 3.10.

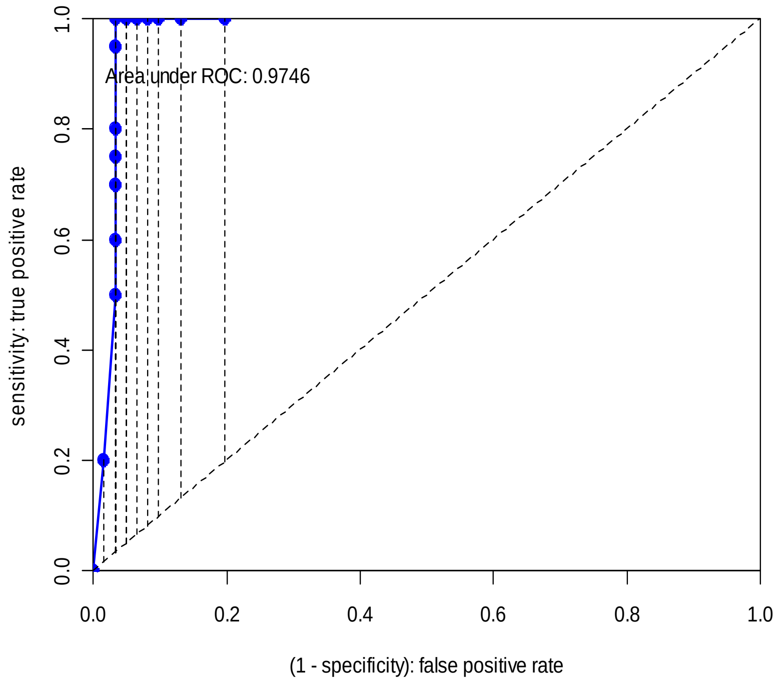


Figure 3.10. ROC for of model-5 with slope, aspect, annual ISR, soil and land cover as predictors

The AUC of the model is excellent (0.98). The ROC curve is far above diagonal, indicating that the prediction is much better than the random one. The model succeeded in fitting the prediction and the observed data. The presence probability map resulting from this model is given in Figure 3.11.

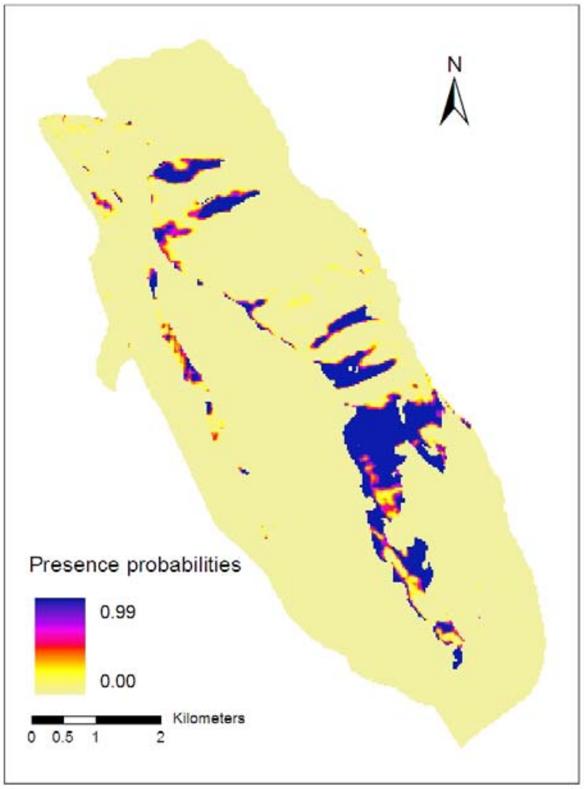


Figure 3.11. Presence probability map of the endemic *Acer* at Mt. Pizzalto resulting from logistic regression model-5

The optimum threshold resulting from the validation was 0.36. By using this threshold, the error matrix from the validation is presented in Table 3.6.

Table 3.6. Error matrix of logistic regression model-5 validation using independent random validation points (n=37)

		Observed	
		Presence	Absence
Predicted	Presence	11	3
	Absence	0	23

Based on data from Table 3.6, some accuracy measures were then calculated with the results summarized in Table 3.7.

Table 3.7. Accuracy measures of logistic regression model-5 validation using independent random validation points (n=37)

Sensitivity	Specificity	Overall accuracy	Kappa statistic
1	0.885	0.919	0.82

Table 3.7 explains the perfect performance of the model in predicting the endemic's presence with the sensitivity value of 1. However, the accuracy in predicting the endemic's absence is lower with the specificity value of 0.885. In overall, 91.9% of observed points can be successfully predicted by the model. The model's Kappa statistic is 0.82. This value indicates the strength of agreement between two data compared to the probability of agreement by chance. Furthermore, by considering the sensitivity and specificity values in some different thresholds, the ROC curve was drawn (Figure 3.12) with the AUC of 0.97.

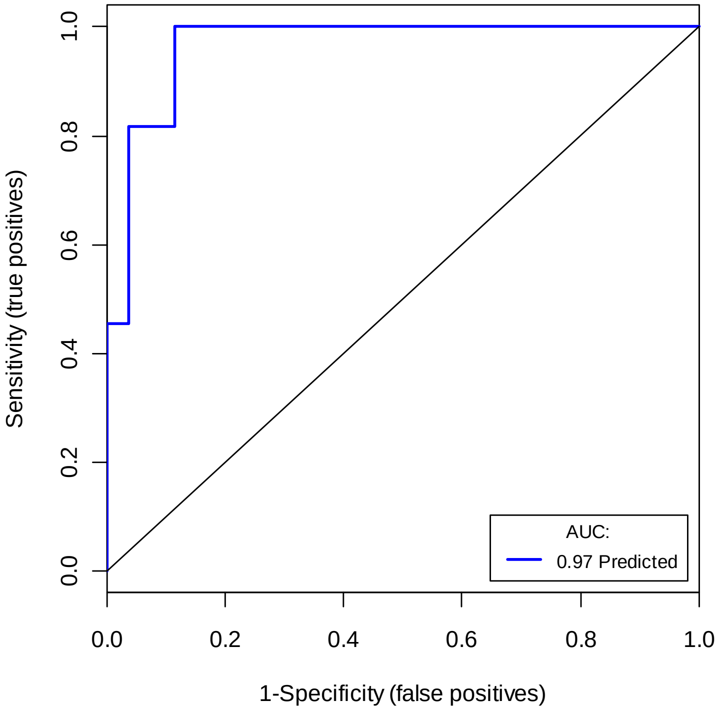


Figure 3.12. ROC of logistic regression model-5 validation using independent random validation points (n=37)

The AUC value from model validation (0.97) is only slightly lower than that from internal validation (0.98). It indicates the robustness of the model. The model succeeded to give a very good accuracy both for training data and validation data. Finally, by using the optimal threshold of 0.36, the predicted presence and absence map of endemic *Acer* resulted from logistic regression model-5 is presented in Figure 3.13.

The logistic regression model-5 predicts the endemic *Acer* potentially occurs in 277.38 ha area of Mt. Pizzalto (6.85 %), spreading in some patches. By excluding the small patches (less than 1 ha), this model predicted 16 patches with an average of 16.5 hectare (Table 3.15)

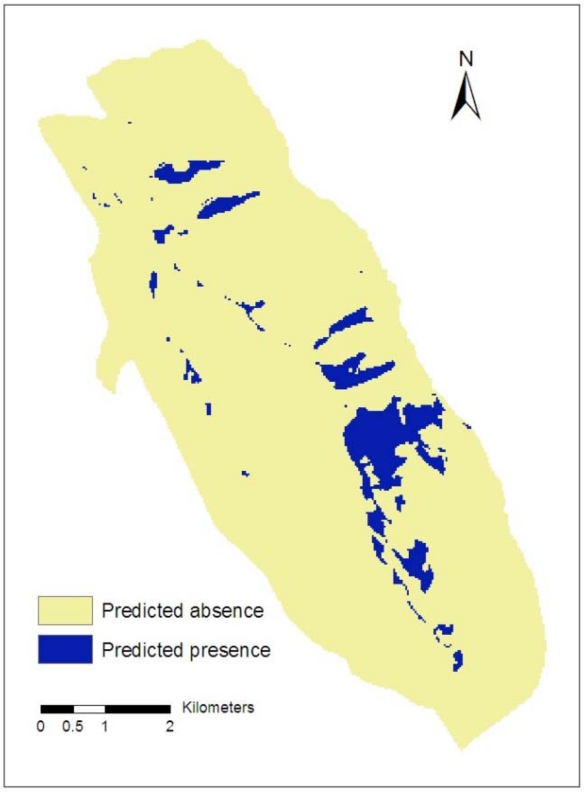


Figure 3.13. Predicted presence and absence map of the endemic *Acer* at Mt. Pizzalto resulting from logistic regression model-5

3.2.2. Geostatistical Modeling

At the first step, a variogram analysis was performed to evaluate the spatial structure of training points. The existence of spatial structure was detected until the distance of about 1100 m (Figure 3.14).

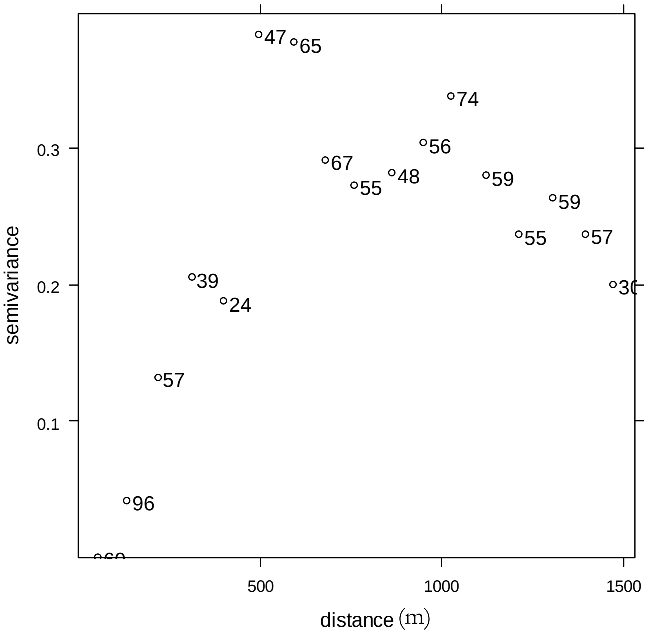


Figure 3.14. Variogram of presence and absence of the endemic *Acer* at Mt. Pizzalto

There is a pattern in which the semivariance value increases following the increase of distance until the distance of 1100 m, although at a distance of 500 m the semivariance values are higher than those in the further distances. This is the answer of the research question 2.b. Based on the variogram, a spherical variogram model was fitted (Figure 3.15) with the parameters as listed in Table 3.8.

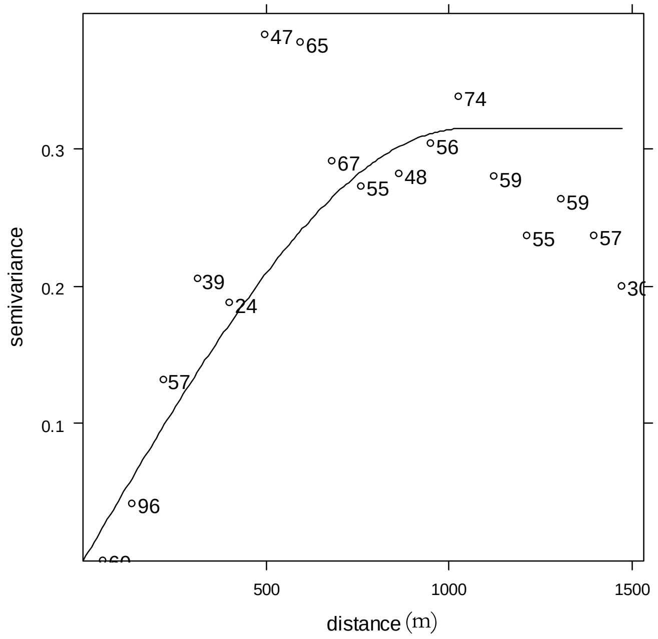


Figure 3.15. Fitted variogram model of presence and absence of endemic *Acer* at Mt. Pizzalto

Table 3.8. Parameters of variogram model of presence and absence of endemic *Acer* at Mt. Pizzalto

Parameter	Value
Nugget	0
Partial sill	0.3145
Range	1040.7

a. Nugget

There is no nugget effect (nugget = 0). This is exactly following the basic theory of variogram in which the semivariance value at the lag of 0 is 0. The empirical variogram even shows that the semivariance value is still 0 at the lag of 88 m. This indicates that there is no variability at distances less than 88 m. At the distances between 0 to 88 m, all pairs consists of presence points (value = 1), leading to the perfect correlation (semivariance = 0).

b. Sill

The sill value is 0.32, indicating the highest semivariance value specified by the variogram model. As spherical model is a bounded one, the semivariance value at the distances of range value (1040 m) or more is exactly 0.32.

c. Range

The range value is 1040. This describes the maximum distance in which spatial structure of the observed distribution of the endemic *Acer* exists. At distances more than 1040 m, there is no more spatial structure (spatial autocorrelation = 0).

The parameters (nugget, sill and range) determine the contribution of sample points in predicting values of unsampled points by interpolation. The parameters were used to assign weights of interpolating points based on their distances to interpolated points. The presence probability value resulting from the interpolation were distributed from -0.13 to 1.06. As the probability should distribute from 0 to 1, the values less than 0 were categorized as 0 and the values more than 1 were categorized as 1. The presence probability map resulting from the kriging interpolation is presented in Figure 3.16.

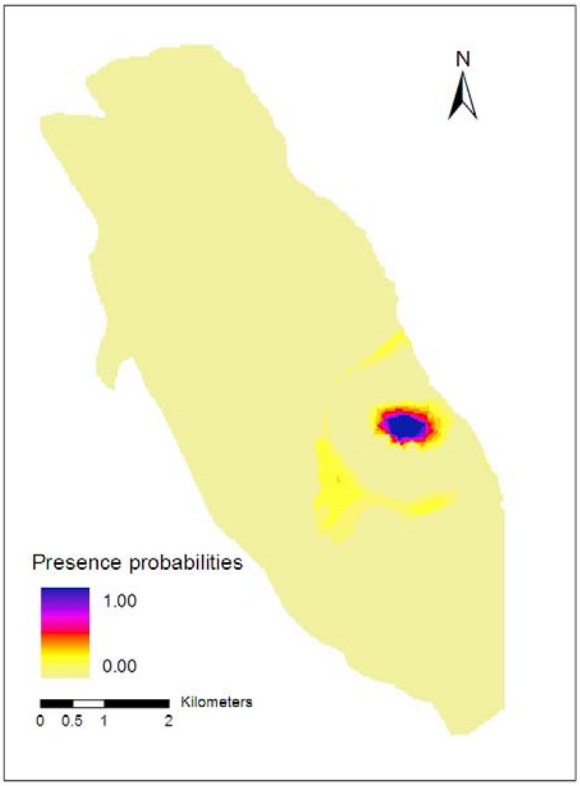


Figure 3.16. Presence probability map of endemic *Acer* at Mt. Pizzalto resulting from geostatistics

There is only one location detected as high probability area (Figure 19). It seems that the geostatistical model succeeded to localize the high probability areas as narrow as the presence training points of the endemic *Acer*. The optimum threshold resulted from the validation was 0.42. By using this threshold, the error matrix from the validation is presented in Table 3.9.

Table 3.9. Error matrix of geostatistical model validation using independent random validation points (n=37)

		Observed	
		Presence	Absence
Predicted	Presence	11	0
	Absence	0	26

Based on data from Table 3.9, accuracy measures were then calculated with results summarized in Table 3.10.

Table 3.10. Accuracy measures of geostatistical model validation using independent random validation points (n=37)

Sensitivity	Specificity	Overall accuracy	Kappa statistic
1	1	1	1

Table 3.10 explains the perfect performance of the model in predicting the endemic's presence and absence. This leads to the perfect value of overall accuracy and Kappa statistic. By using some different thresholds, the sensitivity and specificity values are perfect. This can be seen from the ROC curve (Figure 3.17) with the AUC of 1. This is also the answer of the research question 2.b.

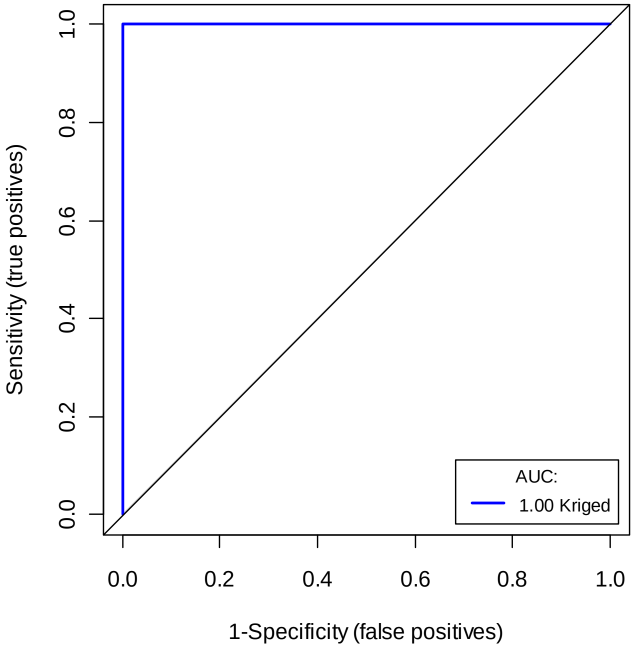


Figure 3.17. ROC of geostatistical model validation using independent random validation points (n=37)

At the final step, by using the optimal threshold of 0.42, the predicted presence and absence map of endemic *Acer* resulting from the geostatistical model is presented in Figure 3.18. The endemic *Acer* was predicted to occur in a patch of 31.68 ha Mt. Pizzalto. This prediction is much lower than that by logistic regression model. Compared to the observed patch of the endemic distribution (5.6 ha), this prediction is 5.7 times higher.

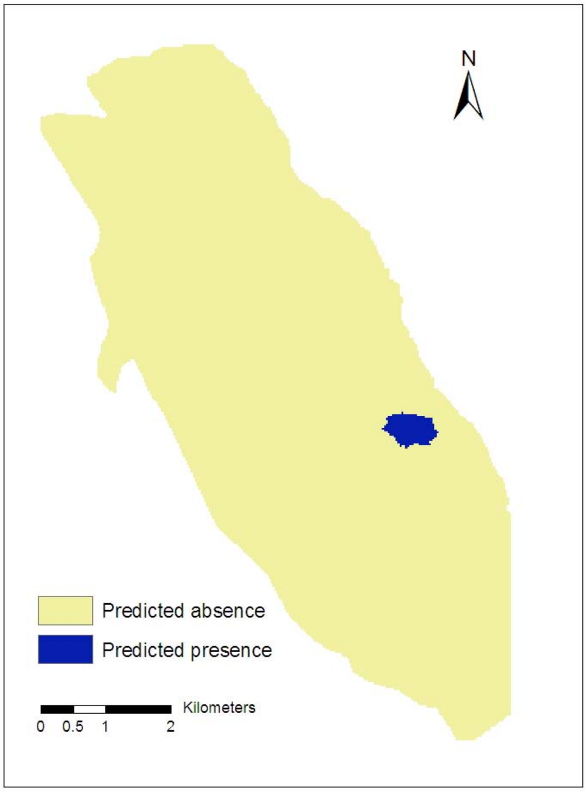


Figure 3.18. Predicted presence and absence map of endemic *Acer* at Mt. Pizzalto resulting from the geostatistical model

3.2.3. Maxent Modeling

The result of jackknife test by using all environmental layers is given in Figure 3.19.

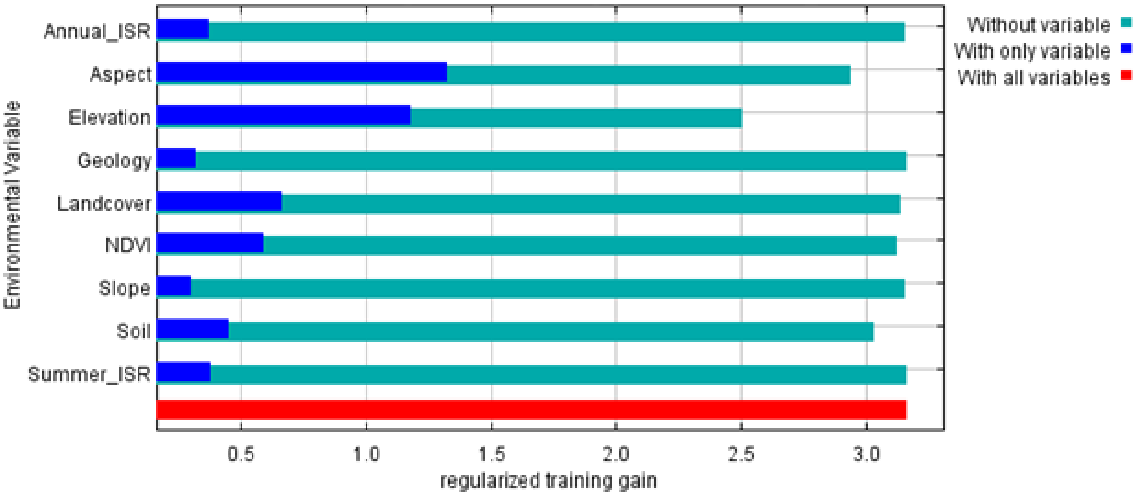


Figure 3.19. Jackknife chart of Maxent modeling of the endemic *Acer*

Figure 3.19 describes the contribution of each variable to the model result both individually and collectively. Elevation is the most contributing variable. Without this variable, the training gain will be the smallest one. On the other hand, summer ISR has the least contribution. There is no different gain by excluding the variable. Based on this result, summer ISR was removed for the next Maxent running. This process was continued until the last variable. The results of the stepwise Maxent are summarized in Table 3.11.

Table 3.11. Results of stepwise Maxent modeling

No	Environmental variables	AUC
1	Elevation, slope, aspect, NDVI, annual ISR, summer ISR, land cover, soil and geology	0.994
2	Elevation, slope, aspect, NDVI, annual ISR, land cover, soil and geology	0.994
3	Elevation, slope, aspect, NDVI, annual ISR, land cover and soil	0.994
4	Elevation, slope, aspect, NDVI, annual ISR and soil	0.994
5	Elevation, aspect, NDVI, annual ISR and soil	0.994
6	Elevation, aspect, annual ISR and soil	0.993
7	Elevation, aspect and soil	0.994
8	Elevation and aspect	0.990
9	Aspect	0.933

Table 3.11 shows the success of Maxent in different number of environmental layers. Model number 1, 2, 3, 4, 5 and 7 have the same accuracy with the AUC of 0.994. By considering the number of environmental variables, model-7 was considered as the best model. This model is the simplest one giving the highest accuracy. Adding more environmental variables give no improvement to the model's accuracy. Model-6 with one more environmental variable even gives a slight lower AUC (0.993). Model-8 was also considered as the best model as removing soil causes just a very slight decrease of AUC (0.004). This means the null hypothesis of the second research hypothesis is rejected.

The influence of each variable of Maxent model-7 can be evaluated using contribution table (Table 3.12), jackknife test result (Figure 3.20) and response curves.

Table 3.12. Relative contributions of the environmental variables to the Maxent model-7

Environmental variables	Percent contributions (%)
Elevation	42.1
Aspect	40.3
Soil	17.6

Figure 3.20. Jackknife test result of Maxent model-7

a. Elevation

Table 3.12 and Figure 3.21 show that elevation is the most contributing variable to the model. Removing this variable leads to the largest reduction of training gain. The relationship between elevation and the presence probability of the endemic *Acer* is not linear (Figure 3.21).

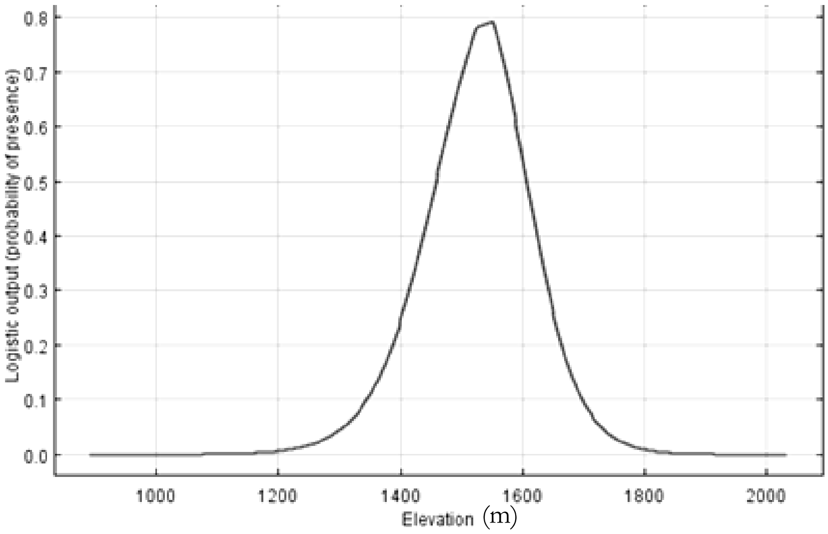


Figure 3.21. Response curve presence probability of the endemic *Acer* at Mt. Pizzalto to elevation

There is no chance for the endemic *Acer* to presence in the elevation below 1200 meter. The presence probability increases exponentially from 1200 meter to about 1500 meter elevation. Between 1500 meter and 1540 meter the presence probability reaches the peak, indicating the highest chances to find this endemic. The probability decreases sharply from 1540 meter to 1800 meter after which there is no chance any more for this endemic to distribute. By using a threshold of 0.5, this endemic potentially occurs in the elevation range of 1450 meter to 1600 meter, *i.e.* in the beech belt (van Gils, *et al.*, 2010).

b. Aspect

The contribution of aspect to the Maxent model-7 is slightly lower than that of elevation. The relationship between this variable and the presence probability resembles a bell shaped curve (Figure 3.22) with the highest presence probability at the aspect of about 130°. By using a threshold of 0.5, the endemic *Ace*r is expected to distribute in areas with the aspect between 80° and 190°, i.e. southern expositions. At northwestern aspects (> 300°), there is no chance to find this endemic.

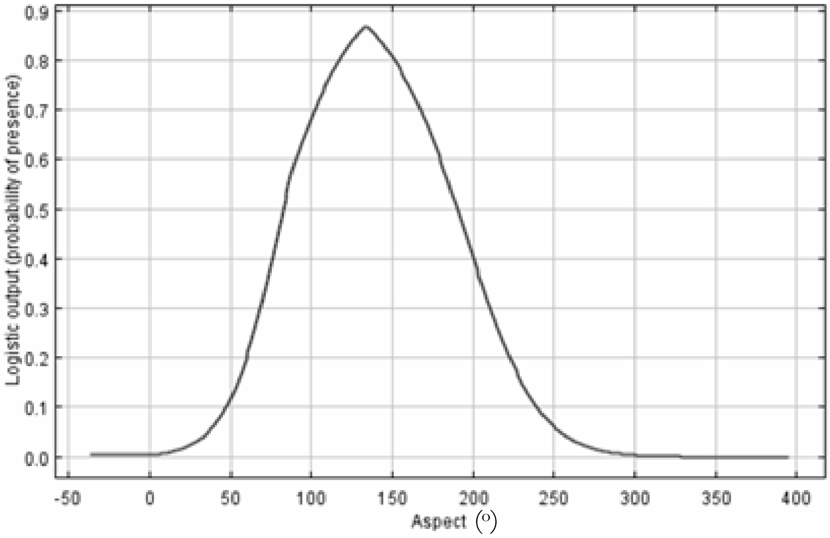


Figure 3.22. Response curve of presence probability of the endemic *Acer* at Mt. Pizzalto to aspect

c. Soil

Soil has the lowest contribution to the Maxent model-7. There is only a little decrease on training gain by removing this variable. Without this variable, there is also only 0.004 decrease of AUC. As soil is categorical, the effect of this variable to the presence probability is also categorical (Figure 3.23). The endemic *Acer* is potentially present only in colluvial soil with the presence probability of about 0.79.

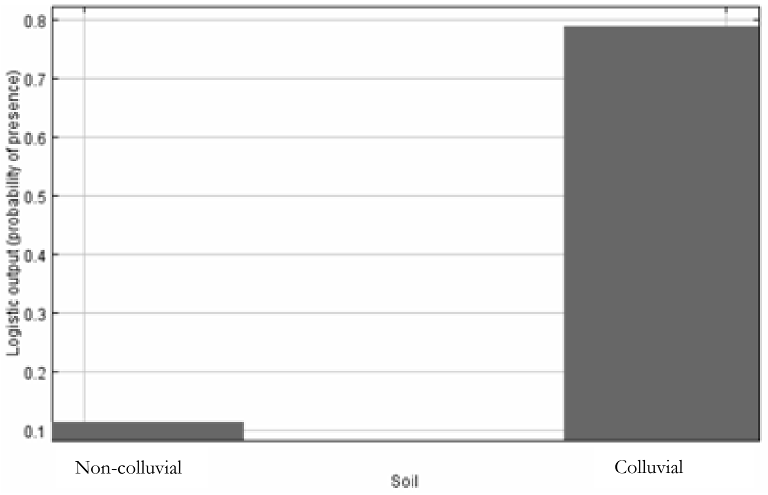


Figure 3.23. Response of presence probability of the endemic *Acer* at Mt. Pizzalto to soil types

The success of Maxent model-7 can be identified from the ROC curve and its AUC (Figure 3.24).

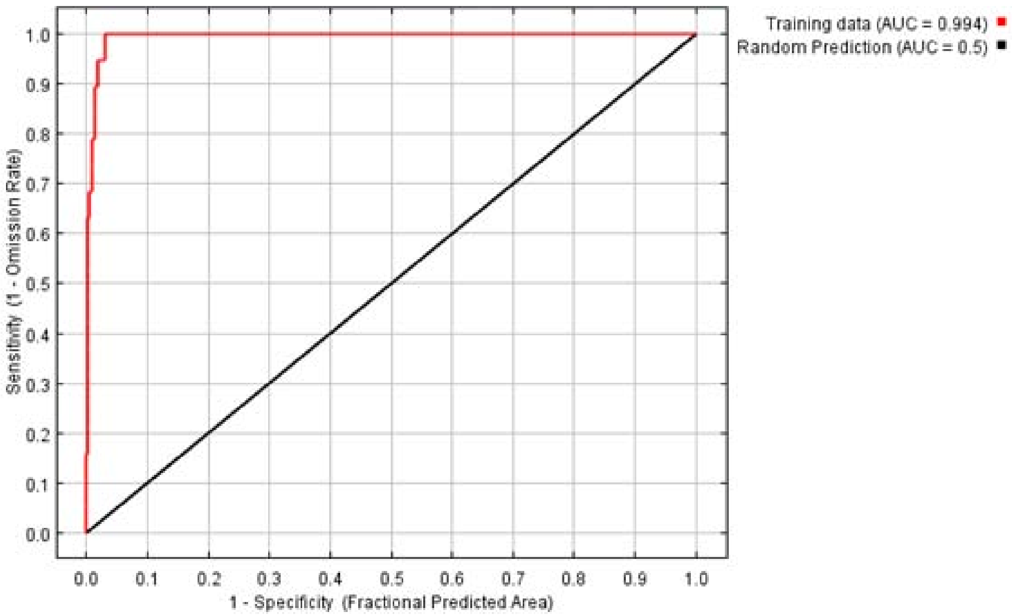


Figure 3.24. ROC for prediction using Maxent model-7

Maxent model-7 was very successful in terms of AUC (0.99). The sensitivity is perfect (1) at many thresholds. This contributes much to bring the ROC curve far above diagonal, indicating that the prediction is much better than the random one. The presence probability map resulting from Maxent model-7 is given in Figure 3.25.

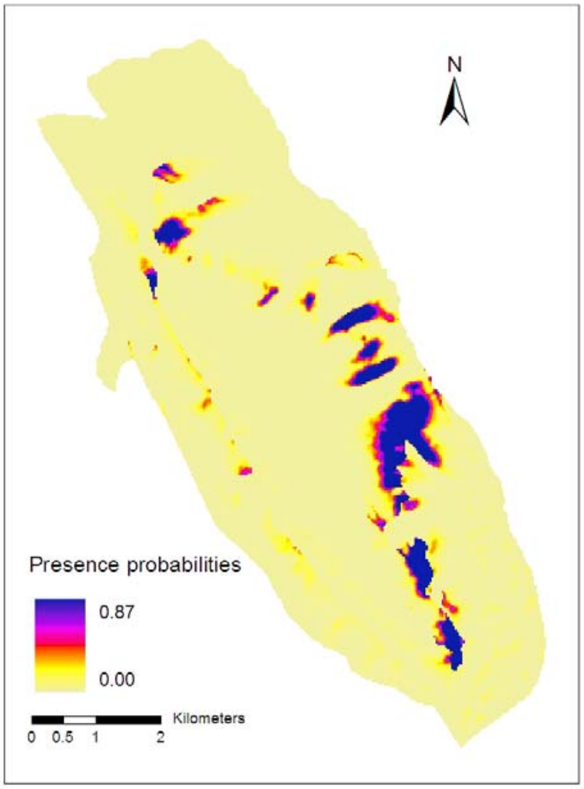


Figure 3.25. Presence probability map of endemic *Acer* at Mt. Pizzalto resulting from Maxent model-7

Different from the previous models (logistic regression and geostatistical models), the highest presence probability resulting from Maxent is 0.87. The distribution pattern of the presence probability is quite similar to that from logistic regression model. Although it is very high, the AUC from Maxent is not comparable to the other models. Since only presence points are used as input, Maxent always generates pseudo absence points for calculating specificity value of ROC. Second validation using independently true presence and absence points is needed for model comparison. The optimal threshold resulted from the validation was 0.275. By using this threshold, the error matrix from the validation is presented in Table 3.13.

Table 3.13. Error matrix of Maxent model-7 validation using independent random validation points (n=37)

		Observed	
		Presence	Absence
Predicted	Presence	11	1
	Absence	0	25

Based on data from Table 3.13, some accuracy measures were then calculated with results summarized in Table 3.14.

Table 3.14. Accuracy measures of Maxent model-7 validation using independent random validation points (n=37)

Sensitivity	Specificity	Overall accuracy	Kappa statistic
1	0.96	0.97	0.94

Table 3.14 shows the perfect success of the model in predicting the endemic's presence with the sensitivity value of 1. For the endemic's absence, the prediction accuracy is slightly lower with the specificity value of 0.96. For both presence and absence of the endemic, Maxent model gave the excellent prediction with the overall accuracy 0.97 and Kappa statistic of 0.94. Furthermore, by considering the sensitivity and specificity values in some different thresholds, the ROC curve was drawn (Figure 3.26) with the AUC of 0.997. Model-8 was also validated with the same procedures and resulting in the same AUC of 0.997. This provides the answer to the research question 2.c. Finally, by using the optimal threshold of 0.275, the predicted presence and absence map of the endemic *Acer* resulting from Maxent is presented in Figure 3.27.

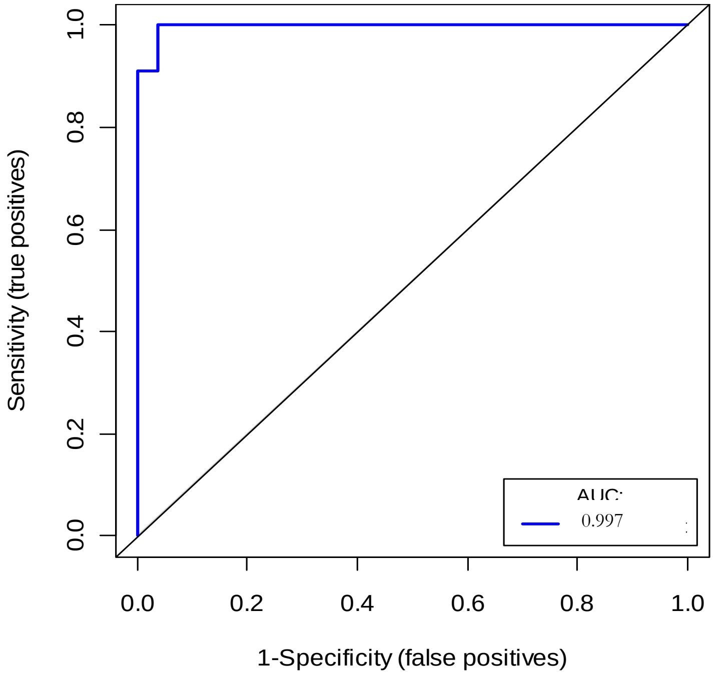


Figure 3.26. ROC of Maxent model-7 validation using independent random validation points (n=37)

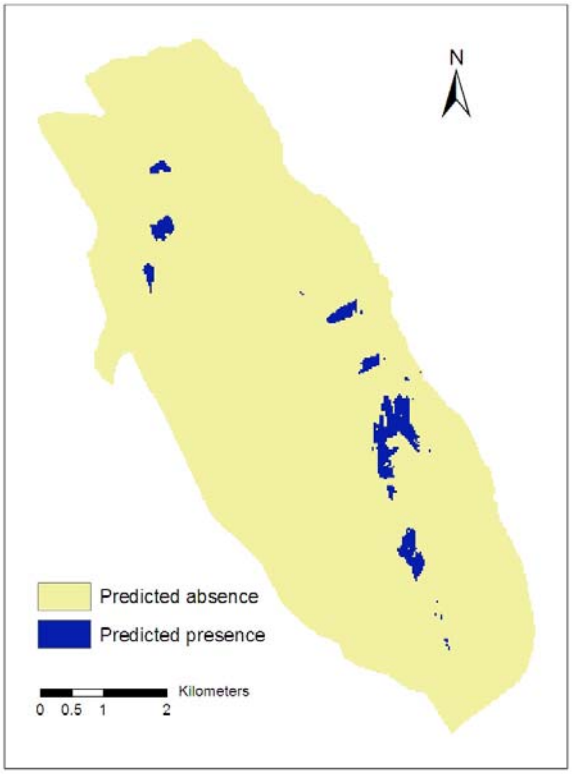


Figure 3.27. Predicted presence and absence map of endemic *Acer* at Mt. Pizzalto resulting from Maxent model-7

The predicted area with the potential occurrence of the endemic *Acer* is 118.26 ha (2.92 % of Mt. Pizzalto area). This predicted area is only 42.63 % of that by logistic regression model. Compared to the prediction by the geostatistical model, the predicted area is 3.73 times higher. This model predicted 8 patches (area > 1 ha) with an average of 12.4 ha (Table 3.15).

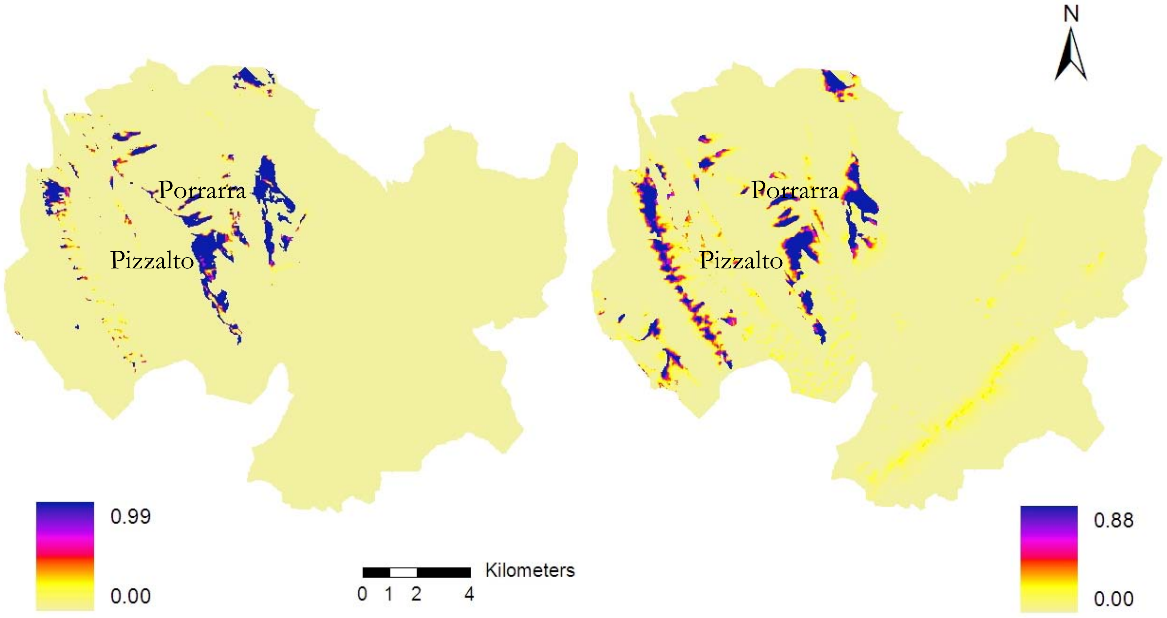
Table 3.15. Patches of predicted distribution of the endemic *Acer* resulting from logistic regression, geostatistics and Maxent (ha); actual patch = 5.6 ha

No	Logistic Regression	Geostatistics	Maxent
1	18.6	31.7	4.5
2	15.1		9.2
3	4.8		4.4
4	3.2		9.0
5	3.6		5.8
6	12.9		46.5
7	4.2		2.1
8	20.0		17.5
9	126.1		
10	2.3		
11	8.4		
12	3.9		

13	1.8		
14	22.2		
15	3.1		
16	3.6		

3.3. Distribution Modeling of The Endemic *Acer* in Southern Majella

Besides at Mt. Pizzalto, the endemic *Acer* is also expected to occur in other places in southern Majella NP. The presence probability of this endemic in southern Majella NP was modeled using the best models from logistic regression and Maxent applied in Pizzalto. The two models were applicable for the extension of prediction since they consider the environmental variables for prediction. Geostatistical model was not used because there were no spatially accurate presence data outside Mt. Pizzalto. The presence probability maps resulting from the two models are shown in Figure 3.28.



A

B

Figure 3.28. Presence probability map of endemic *Acer* in southern Majella NP resulting from logistic regression (A) and Maxent (B)

There is a quite similar pattern of presence probability distribution of the endemic *Acer* in southern Majella NP resulted from logistic regression and Maxent. The two models predicted the high presence probability of this endemic in two mountains, *i.e.* Mt. Pizzalto and Mt. Porrarra. By using a threshold of 0.5, the presence and absence maps resulting from the two models are shown in Figure 3.29.

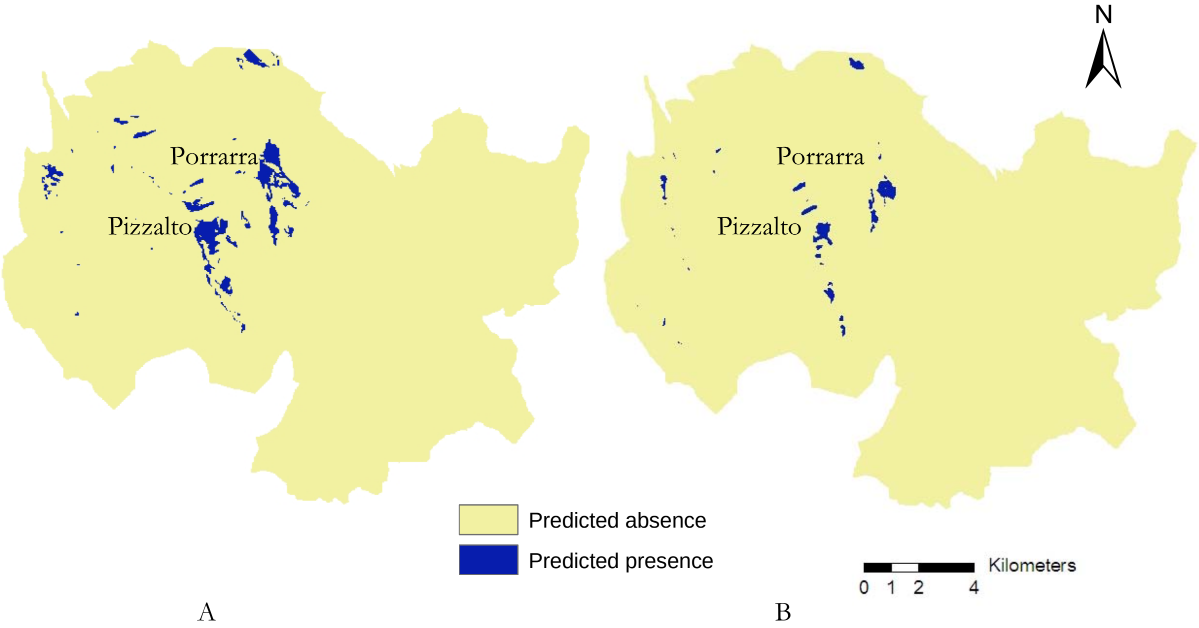


Figure 3.29. Predicted presence and absence map of the endemic *Acer* in southern Majella NP resulting from logistic regression (A) and Maxent (B)

Figure 3.29 describes the difference prediction between logistic regression and Maxent. The predicted presence area resulted from logistic regression is larger than that from Maxent. The logistic regression model predicts the endemic *Acer* potentially distributes in 433 ha area of southern Majella NP, about 2.7 times larger than the prediction by Maxent (162 ha).

4. DISCUSSION

4.1. Distribution Pattern of the Endemic *Acer*

The clumped distribution pattern as found for the endemic *Acer* is common for plant populations (Hutchings, 2003), including tree populations in forests (*e.g.* Frankham *et al.*, 1998). This pattern is also common for rare species. As an example, Chao *et al.* (2008) found that rare trees in rainforest in Taiwan are substantially more clumped than the common ones. The rarer the trees are, the higher the degree of clumping. As a rare species, the endemic *Acer* at Mt. Pizzalto also expresses the same pattern.

Many factors have been related to the clumped distribution pattern of plant population, such as environmental heterogeneity (Dale, 1999), limited distance of seed dispersal, vegetative regeneration (Hutchings, 2003) and species competition (Silvertown & Charlesworth, 2001). The environmental heterogeneity factor assumes that a heterogeneous habitat will provide a mosaic of suitable and unsuitable patches for plant population so that they tend to grow together in the suitable patches. This apparently does not happen to the endemic *Acer* at Mt. Pizzalto. Habitat unsuitability is not the main barrier for the dispersal of this endemic. Beech forest, in which this endemic grows, is widely and homogeneously spread at Mt. Pizzalto. However, the endemic only takes a small area of the beech forest to grow. It means the endemic does not utilize its most suitable habitat.

The more explanatory factor for the clumped distribution of the endemic *Acer* is the vegetative regeneration. Field observation found that most of juveniles of this endemic were produced vegetatively from the roots. Regardless the mortality level (there is no information about this), many juveniles and adult trees of this endemic shows the success of recruitment from vegetative regeneration (Figure 4.1). Obviously, there are short distances between the seedlings and their parent trees as well as among the seedlings themselves. This leads to the clumped distribution.

Figure 4.1. Juveniles (A) and adult trees (B) of the endemic *Acer* growing from vegetative regeneration

The limited distance of seed dispersal can also explain the distribution pattern. The endemic *Acer* can be categorized as an anemochore, in which it exploits wind for transporting fruit and seed (Schulze, *et al.*, 2005). In general, *Acer* trees have winged fruits called samaras (Oterdoom & de Jong, 2001). This will help the fruit to fly and disperse the seeds. However, field observation revealed that the juveniles from germination mostly distributed close to the parent trees. The very dense beech forest is suspected to give barriers for the endemic's fruit to far disperse. In addition, vertically beech trees are dominating the top layer, giving a more difficulty for the endemic's fruits to fly away. This causes most of the endemic's seeds land close to the parent trees. Figure 4.2 shows the endemic *Acer* juveniles which grow from germination. This kind of seedling was mostly found in close distance to the adult trees.

Figure 4.2. The endemic *Acer* juveniles growing from germination

This condition is different from the distance hypothesis developed by Janzen (1970). He provided a hypothesis that an optimum recruitment (high density of new adults) will take distance from parent tree. This distance is equilibrium between seed density and seed survival probability (Figure 4.3). Seed density tends to decrease by the increase of distance from parent tree. However, because of predation, the juvenile mortality is also high at close distance to the parent tree. Therefore he hypothesized that the probability of seed survival increases follows the increase of distance from parent tree.

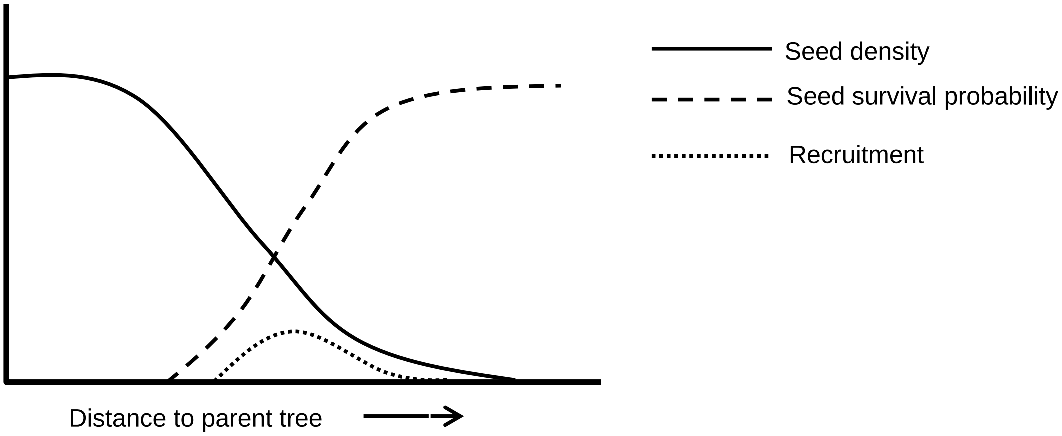


Figure 4.3. Distance hypothesis for new adult recruitment (after Jansen, 1970)

Two arguments can be used to explain why the recruitment pattern of the endemic *Acer* at Mt. Pizzalto is different from the hypothesis, *i.e*:

1. The hypothesis is a general one and was developed from studies on tropical forest trees.
2. The hypothesis on seed survival pattern was based on the juvenile predation.

In tropical forest, there are many predators of tree juveniles such as herbivores and fungi. The endemic *Acer* juvenile at Mt. Pizzalto apparently does not suffer much due to the predation. This maintains the higher seed survival in close distance to parent tree.

Regarding the species competition, clumping is a strategy for reducing interspecific competition (Silvertown & Charlesworth, 2001). This usually happens to population that deals with strong competitors. This also happens to the endemic *Acer* population at Mt. Pizzalto. This population grows in the area in which beech is very dominant. Beech maintains its domination by its shade-tolerance (Piovesan, *et al.*, 2005) and potential allelopathic litter (Hane, *et al.*, 2003). The two strategies enable only beech is recruited under a closed beech canopy with an abundance of beech litter. Consequently, the endemic *Acer* faces a very tough competition. Although the intraspecific competition among the endemic individuals becomes higher, reducing interspecific competition by clumping is preferred for survival.

4.2. Ecological Niches of the Endemic *Acer*

The topographical niches of this endemic at Mt. Pizzalto are summarized in Table 4.1. Land cover, soil and geological types are too broad to be related to the ecological niche.

Table 4.1. Topographical niche of the endemic *Acer* in Pizzalto

No	Environmental factors	Environmental ranges
1	Elevation	1488 – 1590 meter
2	Slope	8^o – 26^o
3	Aspect	84^o - 150^o

Not all areas in the fundamental/potential topographical niches are occupied by the endemic *Acer*. This is because of interspecific competition. The main competitor for the endemic *Acer* is beech which is dominant at Mt. Pizzalto between 1000 and 1800 m. The competition forces the endemic *Acer* narrowing its distribution. This brings the realized niche of this endemic at Mt. Pizzalto is much smaller than its fundamental niche.

Topographical niches of the endemic *Acer* do not show specialization strategy of this endemic. The aspect range of 84^o - 150^o is a common preference for plants growing in the northern hemisphere. This south facing aspect is preferred to get more solar radiation for photosynthesis. The slope range is also categorized as gentle to medium, the generally favourable slope for plant population. The most interesting one is the elevation range. The range is in the middle of the monospecific beech forest, narrower than the general elevation range of this endemic, *i.e.* 750 – 1700 m (Pignatti, 1982). The other beech associated trees such as *Quercus pubescens, Acer obtusatum, Acer platanoides, Acer campestre* and *Acer pseudoplatanus,* tend to specialize in the lower edge of the beech forest (van Gils *et al.*, 2010) to minimize the competition. So, nothing specific is shown by the topographical niches of this endemic. Compared to some studies on European endemics, the endemic *Acer* shows uncommon behaviour. Medail & Verlaque (1997) found that endemic distribution in Southern France and Corsica is positively correlated to specialized habitat and low competition. Casazza, *et al.* (2005) also reported that most of endemics in Italian Maritime Alps grow in communities with the high stress level. Stress tolerant strategy is in line with the habitat specialization and avoiding high competition strategies. On the contrary, the endemic *Acer* even grows in the high interspecific competition area.

How the endemic *Acer* is able to survive under beech domination needs to be more investigated. Although this is uncommon, the existence of rarities with limited distribution in the large and stable forest is possible and might be due to the history of the area (Kruckeberg & Rabinomitz, 1985). In the historical context, it seems the endemic *Acer* habitat was invaded by beech. The fact that *Acer cappadocicum* is a more southern distributed tree and the lack of specialized niche raise a notion that this endemic is a relic of a warm period of Holocene (Giraudi, 2005), about 9,000 to 5,000 years BP (before present). Beech forests were predominant in the Apennines in the cooler period about 2,000 years BP and remains to the present day (Cruise, 1990). Now, this endemic is struggling to defend its existence in its left small area.

4.3. Population Structure of the Endemic *Acer*

The pyramidal shape of the endemic *Acer* population structure (Figure 3.7) describes the ability of this endemic to produce enough new generation to balance the mortality. Juvenile as a baseline of the structure has the highest density. Figure 3.7 also indicates the high mortality in juvenile and sapling stages of the endemic *Acer*. It is common for tree population in nature in which young individuals suffer higher mortality. Juvenile and sapling are the fragile growing stages. The high mortality in these stages can be attributed to diseases, herbivores, disturbances and nutrient deficiencies. According to Yamazaki *et al.* (2008), juvenile mortality in temperate forest was mostly caused by diseases. Figure 4.4 shows damage evidences of the endemic *Acer* juveniles at Mt. Pizzalto that potentially lead to seedling mortality.

Figure 4.4. Evidences of the endemic *Acer* seedling damage: unhealthy (A) and leafless (B)

4.4. Distribution Model Comparison

4.4.1. Accuracy Comparison

All accuracy measures (Table 4.2) identify that geostatistical model is the most successful model. This model has perfect values of sensitivity, specificity, overall accuracy, kappa statistics and AUC. However, the accuracy difference is slight. Compared to outcomes of some studies *e.g.* Loiselle *et al.* (2008), William, et al. (2009) and Guisan, *et al.* (1998), the three models performed extremely well.

Table 4.2. Comparison of model validation results using independent validation points

No	Model	Sensitivity	Specificity	Overall accuracy	Kappa statistics	AUC
1	Logistic regression	1	0.885	0.919	0.82	0.965
2	Maxent	1	0.9615	0.973	0.937	0.997
3	Geostatistics	1	1	1	1	1

The clumped distribution of the observed presence points of the endemic *Acer* may play a role in the model performances. A clumped distribution tends to have spatial structure in which the closer the distance between two points is, the more similar the occurrence status (presence or absence) they have. This enables geostatistics to have an optimal performance. On the contrary, using environmental variables as inputs tends to give prediction for suitable areas, leading to the broader prediction for a clumped distributed population. This affects the logistic regression and Maxent performances in which they have less accurate results. Based on these results, we can say that the distribution of the endemic *Acer* in Pizzalto is slightly better predicted by its spatial structure than by its environmental attributes.

The model accuracy was also be evaluated by using the original predicted probabilities (before classified into presence and absence). Root mean square error (RMSE) of each model was calculated as accuracy measure. Respectively, the RMSE of logistic regression, Maxent and geostatistics are 0.274, 0.203 and 0.109. This also describes that geostatistics is the most accurate one with the lowest error. However, analysis of variance (ANOVA) of the absolute residuals of the three models gave a not significant difference between them with a P value of 0.84. This means that the errors of the three models in predicting distribution of the endemic *Acer* in Pizzalto are statistically the same. Regarding the fourth research hypothesis, the ANOVA result doesn't reject the null hypothesis. This provides the answer to the research question 2.d, either the use of presence and absence points or presence-only points statistically gave the same accuracy.

4.4.2. Predictor Comparison

The best logistic regression model for predicting the endemic *Acer* distribution involves slope, aspect, annual ISR, soil and land cover as explanatory variables. Of the variables, only annual ISR is significant. The overlap distribution of the presence and absence points in the same environmental ranges/classes might cause the insignificant effects of the most variables. Figure 4.5 gives examples of the overlaps for slope and land cover.

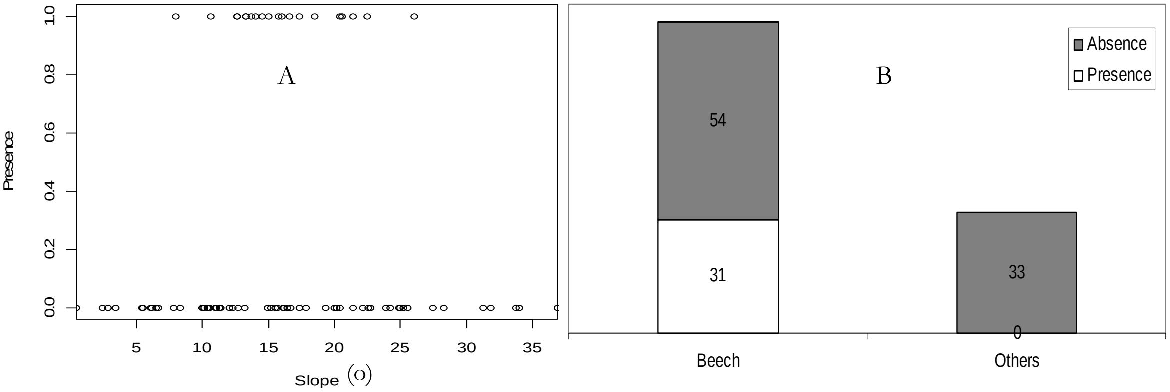


Figure 4.5. Overlap distribution of presence and absence points in slope (A) and land cover (B)

Although most explanatory variables are not significant, they contribute to improve the accuracy. The issue in regression analysis is not that all explanatory variables have to be significant. The issue is on what combination of the variables produces least error. Often, it is a combination of significant and insignificant variables (Rossiter, 2010).

Different from logistic regression, Maxent didn't consider the distribution of absence points of the endemic *Acer*. Maxent learned the distribution of the endemic *Acer* presence points based on the environmental variables, and used the information to specify the presence probability distribution. This enabled the difference of variables of the best model between Maxent and logistic regression. Maxent produced a simple model involving elevation, aspect and soil. We can even say that elevation and aspect are sufficient for the best Maxent since removing soil caused a very slight decrease of AUC (0.004) and gave the same validation AUC of 0.997. So, Maxent needed fewer environmental variables to produce a slightly higher accuracy than the logistic regression did. In this case, the relatively fewer variables needed by Maxent is robust. Table 4.3 shows the robustness in which three environmental variables in different compositions give the excellent AUCs. Slope and annual ISR were involved as substitutes as they relate to elevation and aspect.

Table 4.3. AUC of Maxent with three environmental variables in different compositions

No	Environmental variables	AUC
1	Elevation, aspect and soil	0.994
2	Elevation, annual ISR and soil	0.990
3	Elevation, slope and soil	0.982
4	Slope, aspect and soil	0.974
5	Slope, annual ISR and soil	0.934

Logistic regression and Maxent identified the high influence of topographical factors in predicting the endemic *Acer* distribution. Annual ISR, the only significant variable in logistic regression, relates to aspect and hill shade. For Maxent, elevation and aspect gave 99% contribution to the best prediction. The roles of edaphic factors (soil and geology) and vegetation factors (land cover and NDVI) are not significant. This leads to the notion that the endemic *Acer* distribution at Mt. Pizzalto is actually limited by topographical factors. The significance of annual ISR in logistic regression and the high contribution of aspect in Maxent reveal the influence of solar energy to this endemic distribution. This is in contrast to the Nanyomo (2010) findings. She found aspect and solar radiation provided low contribution to the distribution of five endemics in Majella (*Bunium patraeum, Campanula, Melampyrum italicum, Stipa dasyvaginata* and *Saxifraga italic*). The life form of the endemic *Acer*, *i.e.* tree, is suspected as the main reason of the preference of this endemic to the southern facing areas with high solar radiation.

Different from logistic regression and Maxent, geostatistics didn't consider the environmental variables for prediction. Geostatistics only considers whether the observed presence and absence points spatially correlated or not. It was surprising that geostatistical model gave a perfect prediction to the presence probability of the endemic *Acer* at Mt. Pizzalto as the variogram model doesn't fit very well to the empirical variogram (Figure 3.14). The categorization, *i.e.* all probabilities are finally categorized only into 0 (absence) and 1 (presence), apparently reduced the errors and lifted the accuracy up to 100%. However, since the geostatistical model finally gave a perfect accuracy, this describes two things:

- The existence of spatial structure of the observed presence and absence points of the endemic distribution. This means the null hypothesis of the third research hypothesis, i.e. the endemic *Acer* individuals are randomly distributed, is rejected.
- The success of the variogram model in modeling the spatial structure for kriging interpolation.

4.4.3. Linear versus Non-linear Approaches

Although generalized by logit link function, logistic regression is basically a linear model. It will be optimal when dealing with the linear pattern of presence and absence point distribution on environmental variables, such as more absence points at low elevation and more presence points at high elevation. For all continuous variables, the presence points of the endemic *Acer* don't show linear patterns. This contributed to the insignificances of most variables as well as to more variables needed to achieve the best prediction.

Maxent has different approach. Maxent doesn't give a formula that directly relates environmental variables to presence probability as logistic regression does. Maxent provides response curves to explain the relationship with a nonlinear approach. In defining the probability distribution, Maxent has a constraint that the expected value of each environmental variable under this estimated distribution matches its empirical average. For the endemic *Ace*r data which shows nonlinear pattern, this approach enabled Maxent to provide a better fit between the estimated probability distribution and the empirical value of each variable. This probably also contributed to the fact that two or three environmental variables was sufficient for Maxent to get the best prediction for the endemic *Acer* distribution. As shown in Table 3.11, using aspect alone as environmental variable already gave an excellent AUC (0.933).

As an example, Figure 4.6 presents the different approach of logistic regression and Maxent in providing presence probability distribution of the endemic *Acer* as responses to elevation (as a single variable).

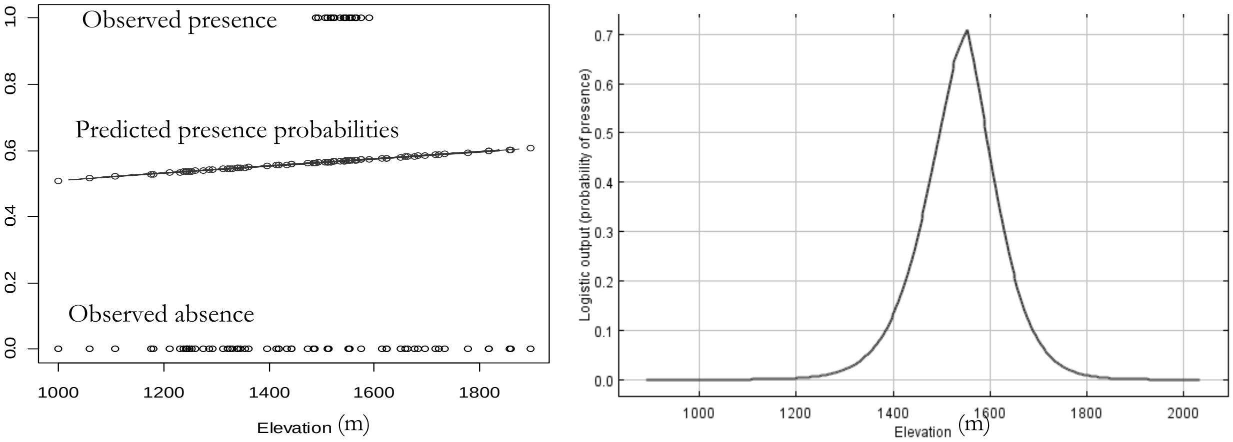


Figure 4.6. Linear approach of logistic regression (A) and nonlinear approach of Maxent (B) in providing probability distribution as responses to elevation

4.4.4. Potential versus Actual distribution

The predicted presence/absence map of the endemic *Acer* resulting from logistic regression (Figure 3.12) indicates that the regression provides the potential distribution of the endemic *Ace*r, not the actual one. Only physical variables were involved in this analysis. In this context, the presence probability is similar to habitat suitability (*e.g.* Pierce & Ferrier, 2000; Keating & Cherry, 2004). Biological variables such as interspecific competition and predation (herbivores and diseases) were not considered. According to Kruckeberg & Rabinomitz (1985), the absence of endemic in its physically suitable environment is usually due to competition from other vegetation, in this case beech.

Similarly to the map generated by logistic regression, the predicted presence/absence map of the endemic *Acer* resulting from Maxent (Figure 3.22) describes the potential distribution of this endemic in Pizzalto. Maxent was actually designed to estimate the species' potential geographic distribution (Philips *et al.*, 2004, Kozak & Wienz, 2006). In reality, species usually don't occupy all of their suitable niches due to some limiting factors such as biotic interactions, barriers for dispersal and human influences (Philips *et al.*, 2006). By assuming that the observed presence points actually have been constrained by such factors, it is possible to model the actual species distribution (realized niche). However, there is usually no proper ecological evidence supporting the assumption (Guisan & Thuiler, 2005). Alternatively, the actual species distribution can be estimated from the modeled distribution by removing some areas that the species is surely known not to inhabit (Philips, 2006).

The predicted patches resulting from logistic regression and Maxent give estimates on area requirement of the endemic *Acer*. The species area requirement is an indication of the minimum area required by a viable population to survive in nature (Leduc & York, 2006). The actual patch size of this endemic, *i.e.* around 5.6 hectare, is in the range of the patch size predicted by logistic regression and Maxent. This indicates that the endemic *Acer* will potentially be capable of sustaining its population for a long period.

Different from the maps resulting from the other two models, the predicted presence/absence map of the endemic *Acer* resulting from indicator kriging (Figure 3.17) indicates the actual distribution of this endemic at Mt. Pizzalto. There is only one spot predicted as distribution area of this endemic. This looks like that geostatistics reconstructed the empirical distribution of the endemic *Acer*. By ignoring the influences of environmental variables, the clumped distribution pattern of this endemic enabled geostatistics to do this. However, this possibly leads to the underestimation if the sample is actually too small.

4.5. Prediction Expansion and Predictive Power Evaluation

Compared to geostatistics, logistic regression and Maxent have an advantage in that the models can be used for expanding the predictions outside the model area. Considering the natural distribution of the endemic *Acer*, the logistic regression and Maxent prediction were expanded to southern Majella NP. This was intended to test the predictive power of the developed models. For this purpose, the predicted presence and absence map of the endemic *Acer* in southern Majella NP resulting from logistic regression and Maxent are displayed together with the distribution of presence points from observation and CRFA geodatabase (Figure 4.7).

The three locations of the CRFA records were visited and no individual of this endemic was found. As the study area is at Mt. Pizzalto, only this mountain was then observed intensively. This endemic was found at Mt. Pizzalto in location with the distance of about 650 meter from the point recorded in the database. With a level-two spatial accuracy, this record has up to two km spatial uncertainty. As shown in Figure 4.7, the observed presence points are inside the range of the uncertainty (buffer), indicating that the two records (CRFA and observed points) can be considered as the same. This means the record in the database indicates the location of the endemic *Acer* population, not an individual record.

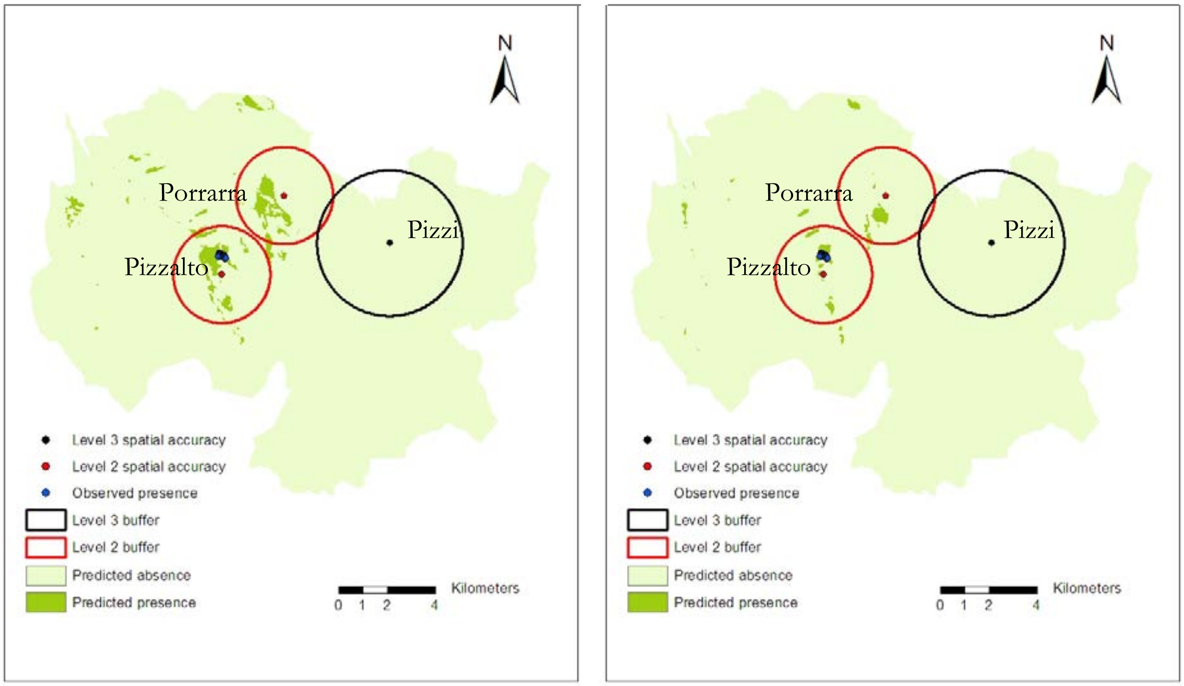


Figure 4.7. Predicted presence and absence map of the endemic *Acer* in southern Majella NP resulting from logistic regression (A) and Maxent (B) and the distribution of presence points from observation and CRFA geodatabase

Logistic regression and Maxent provide a good prediction for this endemic distribution at Mt. Porrarra. Figure 4.7 shows that the patches of this endemic occurrence at Mt. Porrarra predicted by the two models are inside the two km buffer of the CRFA record. As done at Mt. Pizzalto, the predicted patches should be explored to reveal the extent of the actual patch, population density, population structure, distribution and environmental ranges occupied by this endemic. The two models predicted that the endemic *Acer* doesn't distribute at Mt. Pizzi. Although there is a CRFA record at this mountain, we cannot say that the two models gave a wrong prediction since the record itself has a high spatial uncertainty.

In relation to the research question 2.e, Maxent is practically preferable for a broader prediction expansion, such as to central Apennines or southern Italy, because:

- It was shown that the use either presence-only data or present and absence data statistically gave the same accuracy.
- It was also shown that Maxent need fewer environmental variables to produce the same accuracy.
- The most contributing variables for Maxent are elevation and aspect; both of them are DEM derived variables. It is easier to provide such variables compared to soil, geological or land cover map.
- Seventeen records of the endemic *Acer* are available in CRFA geodatabase. However, this study indicates that only records with level-two spatial accuracy (or better) are proper to be used. The records can be used for an initial prediction.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

1. The endemic *Acer* at Mt. Pizzalto showed a clumped distribution pattern with a pyramidal population structure. This endemic was found in a single patch of about 5.6 ha, with the environmental ranges of 1488 -1590 m elevation, 8^o - 26^o slope and 84^o - 150^o aspect. This endemic distributed only in beech forest. Compared to the area of Mt. Pizzalto (4050 ha), locally this endemic can be categorized as a narrow endemic.

2. Slope steepness, aspect, annual ISR, soil type and land cover have been identified by logistic regression as the best combination in explaining the presence probability of the endemic *Acer* at Mt. Pizzalto with a validation AUC of 0.97. The presence probability has a positive relationship with slope and annual ISR. The higher probabilities are also identified in beech forest and colluvial soil type. Of the predictors, only annual ISR is significant. This model predicted 16 patches of the endemic *Acer* potential distributions with an average of 16.5 ha.

3. The endemic *Acer* distribution at Mt. Pizzalto showed a spatial structure, leading to the perfect prediction of the endemic *Acer* occurrences by using indicator kriging with a validation AUC of 1. Only a single patch of 31.68 ha was predicted by this model as distributional area of this endemic.

4. Maxent succeeded in modeling the endemic *Acer* distribution with an almost perfect accuracy (validation AUC of 0.997) using a combination of elevation and aspect as environmental variables. Maxent estimated the high probabilities to find this endemic in the middle elevations (1450-1600 m) and southeastern expositions (80^o-190^o). This model predicted the potential distribution of this endemic in 8 patches with an average of 12.4 ha.

5. The residuals of the three models didn't differ significantly, indicating the same prediction accuracy. The use of presence and absence data was as successful as that of presence-only data.

6. Of the other two CRFA records of the endemic *Acer* in southern Majella NP, only record at Mt. Porrarra is convincing and was justified by logistic regression and Maxent prediction.

5.2. Recommendations

1. The patches predicted by logistic regression and Maxent as potential distribution of the endemic *Acer* at Mt. Pizzalto and Mt. Porrarra should be observed to reveal whether this endemic population is presence or absence.

2. Dealing with a clumped distributed plant, logistic regression, geostatistics and Maxent showed the excellent and perfect prediction. There is a curiosity on the success of the three models if dealing with species with different distribution pattern. This needs to be proven. The more widespread endemics of Majella such as *Campanula fragilis* subs. *cavolinii* and *Centaurea tenoreana* may be selected to be modeled.

3. The observed patch of the endemic *Acer* at Mt. Pizzalto may be developed as a field laboratory for research on this endemic. This area can be called as the endemic *Acer* hotspot. Many questions need to be answered regarding the existence of this endemic at the spot such as on population dynamic, the endemic *Acer* – beech competition, regeneration and expansion or tightening of the spot. The experimental management by removal of beech shadow and litter is also suggested.

4. Maxent is the most promising and efficient model to be applied for distribution modeling of the endemic *Acer* in a wider area. This model needs fewer inputs, *i.e.* presence-only data and fewer environmental variables, to produce the same accuracy. This model may be applied for distribution modeling of this endemic in the Central Apennines (17 records available in CRFA geodatabase) and southern Italy. Elevation and aspect, which are the most contributing variables, are the DEM derived variables that can be easily provided. Since there is no CRFA record with level-one spatial accuracy, the prediction should be done initially in a coarser resolution (*e.g.* 1 km).

LIST OF REFERENCES

Araujo, M.B., & Guisen, A. (2006). Five (or so) challenges for species distribution modeling. *Journal of Biogeography. 33*, 1677 – 1688

Baldwin, R.A. (2009). Use of Maximum Entropy Modeling in Wildlife Research. *Entropy, 11*, 854 – 866.

Bonn, A., Rodrigues, A.S.L., & Gaston, K.J. (2002). Threatened and endemic species: are they good indicators of patterns of biodiversity on a national scale? *Ecology Letters, 5*, 733 – 741.

Brooks, T.M., Mittermeier, R.A., da Fonseca, G.A.B., *et al.* (2006). Global biodiversity conservation priorities. *Science, 313*, 58 – 61.

Caley, K.J. (2008). Biological diversity in Mediterranean Basin. Retrieved 29 July 2010: http://www.eoearth.org/article/Biological_diversity_in_the_Mediterranean_Basin

Casazza, G., Barberis, G., & Minuto, L. (2005). Ecological characteristics and rarity of endemic plants of Italian Maritime Alps. *Biological Conservation, 123*, 361 - 371.

Chang, C. (2010). *Endemic plant distribution in Majella National Park Italy*. Unpublished MSc Thesis, ITC, Enschede.

Chao, W.C., Wu, S.H., Fan, S.W., *et al.* (2008). Distribution pattern of tree species in a lowland rainforest at Nanjen Lake, Sothern Taiwan. *Taiwania, 53* (2), 124 - 133.

Cotgreave, P. and Forseth, I. (2002). *Introductory ecology*. Blackwell Science Ltd, Oxford-London-Edinburgh-Malden-Victoria-Paris.

Cruise, G.M. (1990). Pollen stratigraphy of two Holocene peat sites in the Ligurian Apennines, northern Italy. *Review of Palaeobotany and Palynology, 63*, 299 - 313.

Dale, M.R.T. (1999). *Spatial pattern analysis in plant ecology*. Cambridge University Press, Cambridge-New York-Melbourne-Madrid.

Dickore, W.B., & Miehe, G. (2002). Cold spots in the highest mountain of the world – diversity patterns and gradients in the flora of Karakorum. A chapter book in: *Mountain biodiversity, a global assessment*. Parthenon Publishing, New York – London.

Di Pietro, R. (2009). Observations on the beech woodlands of the Apennines (peninsular Italy): an intricate biogeographical and syntaxonomical issue. *Lazaroza, 30*, 89 - 97.

Elith, J., & Graham, C.H. (2009). Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography, 32*, 66 - 77.

Fox, J. (2002). *Applied regression analysis, linear models and related methods*. Sage, London.

Frankham, R., Ballou, J.D., & Briscoe, D.A. (1998). *A primer of conservation genetics*. Cambridge University Press, Cambridge-New York- Melbourne.

Freeman, E.A., & Moisen, G. (2008). PresenceAbsence: an R package for presence and absence analysis. *Journal of statistical software, 23* (11), 1 - 31.

Freville, H., Colas, B., Riba, M., *et al.* (2004). Spatial and temporal demographic variability in the endemic plant species *Centaurea corymbosa* (Asteraceae). *Ecology, 85* (3), 694 – 703.

Garkoti, S.C. (2008). Estimates of biomass and primary productivity in high altitude maple forest of the West Central Himalayas. *Ecological Research, 23*, 41 - 49.

van Gelderen, D.M., Oterdoom, H.J. & de Jong, P.C. (2001). *Maples of the world.* Timber Press, Inc, Oregon.

Gentry, A.H. (1986). Endemism in tropical versus temperate plant communities. A chapter book in: *Conservation biology, the science of scarcity and diversity*. (Editor: Soule, M.E.). Sinauer Associates Inc, Sunderland – Massachusetts.

van Gils, H.A.M.J., & Loza, A.V. (2006). What drives conversion of tropical forest in Carrasco Provice, Bolovia? *Ambio, 35* (2), 81-85.

van Gils, H., Batsukh, O., Rossiter, D., *et al.* (2008). Forescasting the pattern and pace of Fagus forest expansion Majella National Park, Italy. *Applied Vegetation Science, 11*, 539 – 546.

van Gils, H., Odoi, J.O., & Andrisano, T. (2010). From monospecific to mixed forest after fire?: An early forecast for the montane belt of Majella, Italy. *Forest Ecology and Management, 259* (3), 433 - 439.

Giraudi, C. 2005. Middle to Late Holocene glacial variations, periglacial processes and alluvial sedimentation on the higher Apennine massifs (Italy). *Quaternary Research, 64*, 176 – 184.

Guarino, C., & Napolitano, F. (2006). Community habitats and biodiversity in the Taburno-Camposauro Regional Park. Woodland, rare species, endangered species and their conservation. *Forest@, 3* (4), 527-541.

Guimaraes, R.J.P.S., Freitas, C.J., Dutra, L.V., *et al.* (2009). Spatial distribution of *Biomphalaria* mollusks at Sao Francisco River Basin, Minas Gerais, Brazil, using geostatistical procedures. *Acta Tropica, 109*, 181 - 186.

Guisan, A., Theurillant, J.C., & Kienast, F. (1998). Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science 9,* 65 - 74.

Guisan, A., Weiss, S.B. &. Weiss, A.D. (1999). GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology, 143*, 107-122.

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

Hane, E., Hamburg, S.P., Barber, A.L., & Plaut, J.A. (2003). Phytotoxicity of American beech leachate to sugar maple seedlings in a greenhouse experiment. *Canadian Journal of Forest Research, 33*, 814 – 821.

Hastie, T.J., & Pregibon, D. (1992). Generalized linear model. Chapter 6 of *Statistical Models in S* (editors: Chambers, J.M., & Hastie, T.J.). Wadsworth & Brooks/Cole.

Hernandez, P.A., Franke, I., Herzog, S.K., *et al.* (2008). Predicting species distributions in poorly-studied landscapes. *Biodiversity Conservation, 17*, 1353 - 1366.

Hijmans, R.J. & Graham, C.H. (2006). The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology, 12*, 2272 - 2281.

Hosmer, D.W., & Lemeshow, S. (2000). *Applied logistic regression* (second edition). John Willey and Sons, Inc.

Hutchings, M.J. (2003). The structure of plan populations. A book chapter in: *Plant ecology* (Editor: Crawley, M.J.). Blakwell Publishing, Malden-Oxford-Victoria.

Isaaks, E.H., & Srivastava, R.M. (1989). An Introduction to applied geostatistics. Oxford University Press. New York - Oxford.

Janzen, D.H. (1970). Herbivores and the number of tree species in tropical forests. *American Naturalist, 104*, 501 - 528.

Keating, K.A., & Cherry, S. (2004). Use and interpretation of logistic regression in habitat-selection studies. *Journal of Wildlife Management, 68* (4), 774 - 789.

Kozak, K.H., & Wienz, J.J. (2006). Does niche conservatism promote speciation? A case study in North Americas salamanders. *Evolution, 60* (12), 2604 - 2621.

Kruckeberg, A.R., & Raimowitz, D. (1985). Biological aspects of endemism in higher plants. *Annual Review of Ecology and Systematics, 16*, 447 - 479.

Leduc, E. & York, M.L. (2006). *Biodiversity planning and design: sustainable practices.* Island Press, Washington DC.

Leon-de La Luz, J.L., & Breceda, A. (2006). Using endemic plant species to establish critical habitats in the Sierra de La Laguna Biosphere Reserve, Baja California Sur, Mexico. *Biodiversity and Conservation, 15* (3), 1043 – 1055.

Loiselle, B.A., Jorgensen, P.M., Consiglio, T., *et al.* (2008). Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcome? *Journal of Biogeography, 35*, 105 - 116.

Mackinnon, J., Mackinnon, K., Child, G., & Thorsell, J. (1986). *Managing protected areas in the tropics.* IUCN.

Majella National Park. 2010. *Flora of Majella National Park*. Retrieved 17 August 2010: http://www.parcomajella.it/lgENG/bio_flora.asp

Malcolm, J.R., Liu, C., Neilson, R.P., *et al.* (2005). Global warming and extinction of endemic species from biodiversity hotspots. *Conservation Biology, 20* (2), 538 – 548.

Manel, S., Diaz, J.M, & Ormerod, S.J. (1999). Comparing discriminant analysis and logistic regression for predicting species distributions: a case study with Himalayan river bird. *Ecological Modelling, 120* (2-3), 337 – 347.

Mauseth, J.D. (2009). *Botany: an introduction to plant ecology* (fourth edition). Jones and Barlett Publishers, Sudbury-Ontario-London.

Medail, F., & Veralque, R. (1997). Ecological characteristics and rarity of endemic plats from southeast France and Corsica: implications for biodiversity conservation. *Biological Conservation, 80*, 269 – 281.

Meffe, G.K., & Carol, C.R. (1994). *Principles of Conservation Biology*. Sinauer Associates Inc, Sunderland – Massachusetts.

Miller, J. (2005). Incorporating spatial dependence in predictive vegetation models: residual interpolation method. *The Professional Geographer, 57* (2), 169 - 184.

Mittermeier, R.A., Myers, N., Thomsen, J.B., *et al.* (1998). Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. *Conservation Biology, 12*, 516 – 520.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., *et al.* (2000). Biodiversity hotspots for conservation priorities. *Nature, 403*, 853 – 858.

Myers, N. (2001). Hotspots. *Encyclopedia of Biodiversity*. 371 – 381.

Nanyomo, S. 2010. *Spatial modeling of endemic plants hotspots: a case of in Majella National Park in Italy.* Unpublished MSc Thesis, ITC, Enschede.

Odum, E.P. (1971). *Fundamentals of ecology* (third edition). W.B. saunders company, Philadelphia-London-Toronto.

Oterdoom, H.J., & de Jong, P.C. (2001). Structure of Maples. A book chapter in: *Maples of the world* (Editor: van Gelderen, D.M., Oterdoom, H.J., & de Jong, P.C.). Timber Press, Inc. Oregon.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences, 30* (7), 683 - 691.

Peterson, A.T., & Navarro-Siguenza, A.G. (1999). Alternate species concepts as bases for determining priority conservation area. *Conservation Biology, 13*, 427 – 431.

Philips, S.J., Dudik, M., & Schapire, R.E. (2004). A Maximum Entropy approach to species distribution modeling. *Proceedings of the 24th International Conference on Machine Learning (ICML).* New York.

Philips, S.J., Robert P. Anderson and Robert E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modeling 190: 231 - 259.

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

Philips, S. (2010). A brief tutorial on Maxent. Retrieved 10 October 2010:

http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc

Pierce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modeling, 133*, 225 - 245.

Pignatti, S. (1982). *Flora D'Italia* (volume secundo). Edagricole, Bologna.

Piovesan, G., Di Filippo, A., Alessandrini, A., *et al.* (2005). Structure, dynamics and dendroecology of an old-growth Fagus forest in the Apennines. *Journal of Vegetation Science*, 16, 13 – 28.

Prasad, S.N., Vijayan, L., Balachandran, S., *et al.* (1998). Conservation planning for the Western Ghats of Kerala: A GIS approach for location of biodiversity hot spots. *Current Science, 75* (3), 211 – 219.

Primack, B.C. (1993). *Essential of conservation biology*. Sinauer Associate Inc. Sunderland – Massachusetts.

Reid, W.F. (1998). Biodiversity hotspots. *Tree, 13*, 275 - 280.

Robertson, G.P. (1987). Geostatistics in ecology: interpolating with known variance. *Ecology, 68* (3), 744 - 748.

Rossiter, D.G. & Loza, A. (2010). Technical note: Analyzing land cover change with logistic regression in R. Retrieved 14 July 2010: http://www.itc.nl/personal/rossiter/teach/R/R_lcc.pdf

Rossiter, D.G. (personal communication, October 25, 2010).

Rushton, S.P., Ormerod, S.J. & Kerby, G. (2004). New paradigm for modeling species distributions? *Journal of Applied Ecology 41*, 193-200.

Schulze, E.D., Beck, E., & Muller-Hohenstein, K. (2005). *Plant ecology*. Springer, Berlin-Heidelberg.

Shi, H., Singh, H., Kant, S., *et al.* (2005). Integrating habitat status, human population pressure, and protection status into biodiversity conservation priority setting. *Coservation Biology. 19* (4), 1273 – 1285.

Silvertown, J.W., & Charlesworth, D. (2001). *Introduction to plant biology*. Blackwell Science Ltd, Oxford-London-Edinburg-Malden-Victoria –Paris.

Spector, C. (2010). Soil forming factors. Retrieved 9 November 2010: http://soils.gsfc.nasa.gov/soilform/deposits.htm.

Takezawa, K. (2006). *Introduction to nonparametric regression*. John Wiley and Sons, Inc., New Jersey.

Thompson, J.D. (2005). *Plant evolution in the Mediterranean*. Oxford University Press, Oxford.

Thcouto, M.G.P., Yemefack, M., Deboer, W.F., *et al.* 2006. Biodiversity hotspots and conservation priorities in the Campo-Ma'an rain forests, Cameroon. *Biodiversity and Conservation, 15* (4), 1219 – 1252.

Tribsch, A., Englisch, T., Gugerli, F., *et al.* (2010). Integrating data across biodiversity levels. A chapter book in: *Data mining for global trends in mountain biodiversity.* (Editor: Spehn, E.M., & Korner, C.). CRC Press, Boca Raton.

Veetas, O.R., & Grytnes, J.A. (2002). Distribution of vascular plant species richness and endemic richness along the Himalayan elevation gradient in Nepal. *Global Ecology and Biogeography, II*, 291 – 301.

Watts, D. (1984). The spatial dimension of biogeography. A chapter book in: *Biogeography, recent advances and future directions.* (Editor: Taylor, J.A.). Barnes and Noble Book, New Jersey.

Webster, R., & Oliver, M.A. (2007). *Geostatistics for environmental scientist* (second edition). John Wiley and sons Ltd, Cichester.

Wikipedia. 2010. *Flora of Italy*. Retrieved 30 July 2010: http://en.wikipedia.org/wiki/Flora_of_Italy

Williams, J.N., Seo, C., Thorne, J., *et al.* (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions, 15*, 565 – 576.

Wilson, K.A., McBride, M.F., Bode, M., & Possingham, H.P. (2006). Prioritizing global conservation efforts. *Nature, 440*, 337 – 340.

Yamazaki, M., Iwamoto, S., & Seiwa, K. (2008). Distance- and density-dependent of seedling mortality causes by several diseases in eight tree species co-occuring in a temperate forest. *Plant Ecology, 201* (1), 181-196. A book chapter in: *Forest ecology: recent advances in forest ecology* (editor: van der Valk, A.). Springer.

APPENDICES

Appendix 1. Observed presence points of the endemic *Acer* in Pizzalto

No	X	Y	Number of juveniles	Number of saplings	Number of poles	Number of trees
1	423870	4643120	11	6	0	0
2	423852	4643090	1	0	1	1
3	423867	4643110	21	12	0	0
4	423893	4643110	4	6	4	1
5	423860	4643060	4	18	0	0
6	423905	4643070	0	4	0	0
7	423917	4643070	5	2	0	0
8	423894	4643050	1	1	0	0
9	423928	4643050	26	5	0	0
10	423954	4643030	14	2	0	1
11	423962	4643010	38	12	11	5
12	423978	4643020	4	9	0	0
13	424028	4643000	6	6	0	0
14	424044	4643020	0	1	0	0
15	424038	4643030	7	9	0	0
16	423995	4642990	2	5	1	0
17	423962	4642970	3	7	0	0
18	423936	4642990	4	3	0	0
19	423900	4642980	0	9	2	0
20	423865	4642970	4	6	0	0
21	423861	4643000	1	1	0	0
22	423843	4642980	7	7	1	0
23	423842	4643000	29	12	2	0
24	423817	4642970	5	5	0	0
25	424031	4642980	42	36	1	0
26	424045	4642980	61	30	3	0
27	424040	4642950	31	3	5	1
28	424046	4642920	10	16	1	0
29	424086	4642900	2	4	0	0
30	424089	4642870	0	1	0	0
31	424134	4642890	26	7	3	0

Notes:

1. Coordinate projection system: WGS 1984 UTM Zone 33 N
2. Juveniles: new generation with height < 1.5 m.
3. Sapling: young tree with height > 1.5 m and DBH < 10 cm.
4. Pole: tree with DBH 10 – 20 cm
5. Tree: tree with DBH > 20 cm

Appendix 2. Observed absence points of the endemic *Acer* in Pizzalto

X	Y	Land cover
423561	4642637	Beech forest
419679	4644647	Beech forest
421725	4642788	Beech forest
422152	4642872	Beech forest
421201	4642723	Beech forest
419404	4644176	Beech forest
423139	4642695	Beech forest
424118	4642798	Beech forest
424478	4642208	Beech forest
421246	4646431	Beech forest
422107	4645587	Beech forest
422085	4645331	Beech forest
421970	4646275	Beech forest
422966	4643473	Beech forest
423066	4643601	Beech forest
423235	4643706	Beech forest
423518	4643845	Beech forest
423820	4643985	Beech forest
419768	4646642	Beech forest
419691	4647056	Beech forest
419515	4647512	Beech forest
420571	4646592	Beech forest
420622	4646858	Beech forest
419750	4647772	Beech forest
425016	4640654	Beech forest
424849	4640578	Beech forest
424503	4640457	Beech forest
424239	4640350	Beech forest
424143	4640286	Beech forest
423751	4642898	Beech forest
423543	4641665	Beech forest
423766	4641679	Beech forest
420571	4647783	Beech forest
420260	4644627	Beech forest
420766	4644212	Beech forest
421375	4644015	Beech forest
421355	4644316	Beech forest
423290	4645463	Beech forest
422676	4645502	Beech forest
422611	4645114	Beech forest
422531	4644856	Beech forest
422495	4644633	Beech forest
422434	4644538	Beech forest
423659	4643444	Beech forest

423286	4642478	Beech forest
423485	4642508	Beech forest
421718	4642791	Beech forest
419082	4647372	Beech forest
419893	4646361	Beech forest
425000	4643042	Beech forest
423473	4643345	Beech forest
424186	4643487	Beech forest
424343	4643513	Beech forest
418353	4647813	Beech forest
425317	4641979	Cereal
425093	4641979	Cereal
425318	4640742	Cereal
425565	4640928	Cereal
421183	4641861	Farmland
424897	4638269	Farmland
419674	4644504	Grassland
419968	4644579	Grassland
421326	4642207	Grassland
421413	4642632	Grassland
424698	4642144	Grassland
424908	4641068	Grassland
423973	4642049	Grassland
420787	4648985	Grassland
420760	4648480	Grassland
419953	4646081	Grassland
419449	4643661	Grassland
419996	4643923	Grassland
420202	4644066	Grassland
424049	4644221	Grassland
424191	4642437	Grassland
424739	4642903	Grassland
423347	4643305	Grassland
422008	4646933	Grassland
419074	4647850	Grassland
421561	4642619	Grassland (little stone)
423207	4643289	Grassland (scattered juniper)
423885	4643456	Shrubland (small juniper)
422995	4642782	Stony grassland
421888	4642786	Stony grassland
422669	4643638	Stony grassland
422305	4643644	Stony grassland
422143	4642923	Stony grassland (scattered *Pinus nigra*)

Notes:
Coordinate projection system: WGS 1984 UTM Zone 33 N

Appendix 3. "R" scripts

1. logistic regression analysis

```
# Reading the data
> acer=read.csv("training_regression.csv")
> str(acer)
> attach(acer)

# Stepwise logistic regression
> glm.endemic1= glm(Presence ~
Elevation+Slope+Aspect+NDVI+Annualisr+Geology+Soil+Landcover, family = binomial,data
= acer)
> summary(glm.endemic1)

> glm.endemic2= glm(Presence ~ Elevation+Slope+Aspect+NDVI+Annualisr+Soil+Landcover,
family = binomial,data = acer)
> summary(glm.endemic2)

> glm.endemic3= glm(Presence ~ Elevation+Slope+Aspect+NDVI+Annualisr+Soil, family =
binomial,data = acer)
> summary(glm.endemic3)

> glm.endemic4= glm(Presence ~ Slope+Aspect+NDVI+Annualisr+Soil, family =
binomial,data = acer)
> summary(glm.endemic4)

> glm.endemic5= glm(Presence ~ Slope+Aspect+NDVI+Annualisr+Soil+Landcover, family =
binomial,data = acer)
> summary(glm.endemic5)

> glm.endemic6= glm(Presence ~ Slope+Aspect+Annualisr+Soil+Landcover, family =
binomial,data = acer)
> summary(glm.endemic6)

> glm.endemic7= glm(Presence ~ Slope+Aspect+Annualisr+Soil, family = binomial,data =
acer)
> summary(glm.endemic7)

> glm.endemic8= glm(Presence ~ Slope+Annualisr+Soil+Landcover, family =
binomial,data = acer)
> summary(glm.endemic8)

> glm.endemic9= glm(Presence ~ Slope+Annualisr+Soil, family = binomial,data = acer)
> summary(glm.endemic9)

> glm.endemic10= glm(Presence ~ Annualisr+Soil, family = binomial,data = acer)
> summary(glm.endemic10)
```

```
# Accuracy assessment for the best model with the lowest AIC, i.e. glm.endemic6
> d2 = function(model) {round(1 - (model$deviance/model$null.deviance),4)}
> d2(glm.endemic6)
> source("lcc.R")
> logit.plot.quad(glm.endemic6)
> r = logit.roc(glm.endemic6)
> logit.roc.area(r)
> logit.roc.plot(r, "ROC for prediction by slope, aspect, annualisr, land cover and
soil")
```

2. Indicator kriging analysis

```
# Data reading
> library(gstat)
> a=read.csv("training_kriging.csv")
> str(a)

# Converting data frame to spatial data frame
> class(a)
> coordinates(a)=c("X","Y")
> class(a)

# Sample variogram analysis
> v1=variogram(Presence~1,data=a)
> plot(v1,pl=T)
> v2=variogram(Presence~1,data=a,cutoff=1500,width=100)
> plot(v2,pl=T)
> v3=variogram(Presence~1,data=a,cutoff=1500,width=150)
> plot(v3,pl=T)
> v4=variogram(Presence~1,data=a,cutoff=1500,width=50)
> plot(v4,pl=T)
> v5=variogram(Presence~1,data=a,cutoff=1500,width=75)
> plot(v5,pl=T)
> v6=variogram(Presence~1,data=a,cutoff=1500,width=90)
> plot(v6,pl=T)

# Fitting variogram model for the best sample variogram, i.e. v6
> vm=vgm(0.3,"Sph",1000,0)
> vmf=fit.variogram(v6,vm)
> plot(v6,model=vmf,pl=T)
> vmf
```

Note: kriging interpolation was done in ArcMap by using the parameters resulted from the fitted variogram model.

3. Logistic regression model validation

```
# Data reading
> library(PresenceAbsence)
> val1=read.csv("test_regression.csv")
> str(val1)

# Identifying optimal threshold
> optimal.thresholds(val1)

# Accuracy assessment at the selected optimal threshold, i.e. 0.36
> cm0.36=cmx(val1,threshold=0.36)
> cm0.36
> sensitivity(cm0.36,st.dev=F)
> specificity(cm0.36,st.dev=F)
> Kappa(cm0.36,st.dev=F)
> auc(val1,st.dev=F)
> auc.roc.plot(val1,col="blue")
> abline(0,1,lty=1)
```

4. Geostatistical model validation

```
# Data reading
> library(PresenceAbsence)
> val2=read.csv("test_kriging.csv")
> str(val2)

# Identifying optimal threshold
> optimal.thresholds(val2)

# Accuracy assessment at the selected optimal threshold, i.e. 0.42
> cm0.42=cmx(val2,threshold=0.42)
> cm0.42
> sensitivity(cm0.42,st.dev=F)
> specificity(cm0.42,st.dev=F)
> Kappa(cm0.42,st.dev=F)
> auc(val2,st.dev=F)
> auc.roc.plot(val2,col="blue")
> abline(0,1,lty=1)
```

4. Maxent model validation

```
# Data reading
> library(PresenceAbsence)
> val3=read.csv("test_maxent.csv")
> str(val3)
```

```
# Identifying optimal threshold
> optimal.thresholds(val3)

# Accuracy assessment at the selected optimal threshold, i.e. 0.23
> cm0.23=cmx(val3,threshold=0.23)
> cm0.23
> sensitivity(cm0.23,st.dev=F)
> specificity(cm0.23,st.dev=F)
> Kappa(cm0.23,st.dev=F)
> auc(val3,st.dev=F)
> auc.roc.plot(val3,col="blue")
> abline(0,1,lty=1)
```