University of Tartu
Faculty of Science and Technology
Institute of Ecology and Earth Sciences
Department of Geography

Master's thesis in Geoinformatics for Urbanised Society (30 ECTS)

Suitability Analysis for Alvars in Estonia using Random Forest and GIS based Multi Criteria Decision Making approach

Irada Ismayilova

Supervisor: Evelyn Uuemaa, PhD
Co-supervisor: Aveliina Helm, PhD,
Department of Botany,
University of Tartu

Allowed to defence:
Supervisor:
Head of Department

Tartu 2020

Abstract

Alvars are one of the most species rich habitats in Estonia. Anthropogenic pressure in the form of land use change has unwanted consequences on the grassland's persistence. Therefore, their conservation and restoration issue is becoming more and more relevant. Many attempts for their restoration have already been made. However, land suitability analysis, using two different techniques, was performed for the first time in this thesis. As such, Random Forest (RF) method of Machine Learning technique and Land suitability analysis, together with Multi Criteria Decision Making (MCDM) approach was utilized. RF predicted 610.91 km^2 while MCDM method predicted 987.93 km^2 of suitable areas for alvar restoration or creation of alvar-like habitats in Estonia. Results of suitability analysis might later be used by decision makers in future alvar restoration works.

Key words: Alvars, land suitability, analytic hierarchy process, weighted overlay analysis, random forest

CERCCS code: P510- Physical geography

Sisututvustus

Alvarid on üks liigirikkamaid elupaigatüüpe Eestis. Inimtekkeline surve maakasutuse muutumise näol on avaldanud rohumaade püsimajäämisele soovimatuid tagajärgi. Seetõttu on nende säilitamise ja taastamise küsimus üha aktuaalsem. Nende taastamiseks on tehtud juba palju katseid. Käesolevas töös aga viidi esmakordselt läbi sobivusanalüüs, kasutades kaht erinevat tehnikat. Kasutati masinõppe meetoditest otsustusmetsa (ingl lühend RF) ja sobivus analüüsi koos mitmekriteeriumilise otsustusanalüüsi (ingl lühend MCDM) meetodiga. Otsustusmetsa meetodiga prognoositi, et loopealsete või loopealselaadsete elupaikade taastamiseks sobivad alasid on Eestis 610,91 km^2, MCDM-meetod aga andis tulemuseks 987,93 km^2. Sobivusanalüüsi tulemusi on võimalik otsustajatel edaspidi kasutada alvarite taastamistööde käigus.

Võtmesõnad: alvarid, maa kasutusotstarve, analüütiline hierarhiline otsustusprotsess, kaardialgebra, otsustusmets

CERCCS kood: P510 – Loodusgeograafia

Table of Contents

1. Introduction

Alvar grasslands are calcareous habitats that can be found in Estonia, Sweden and in few other places on the Northern hemisphere in a limited quantity. Alvar grasslands are of immense importance due to their species richness, variety of important ecosystem services that they provide and because they hold natural and cultural heritage in European Landscapes. In Estonia, alvars are the type of grasslands that were developed under human influence, especially due to grazing practices. However, with the change of land use during the past century, existing grasslands became overgrown with shrubs and trees and their areal distribution dramatically decreased, leading also to decline of area of suitable habitat for many species related to these grasslands. Fragmented grassland patches also scored the process of alvar disappearance. Further, alvars can be very different from each other depending on the environment they are exposed to. Considering the high value of alvar grasslands and being priority habitat type in Natura 2000 (Eriksson & Rosén, 2008), these grasslands need restoration and conservation and not only within the territory of Estonia. Many studies had been carried out in order to study history of the grasslands, species composition and plant diversity of the alvars, also the response of those species communities on habitat loss (Helm & Pärtel, 2006). Also large restoration activities have been carried out in order to restore existing alvars by conducting clean-up works and removing unwanted plants such as Juniper (Juniperus communis) and scots pine (Pinus sylvestris). For example, ca 3000 ha of overgrown alvar grasslands were restored in Western Estonia from 2014 to 2019 during the project LIFE to Alvars (LIFE to alvars).

Land/habitat suitability analysis is one of the most frequently used techniques in environmental management. As a result of a land use change and resulting habitat fragmentation and loss, necessity to find alternative and/or most suitable lands for restoration and conservation has increased. There are many examples where land suitability analysis has been applied for planning habitat restoration. For example, Novak and Short (2000) performed suitability analysis for eelgrass meadows in Plum Island and Hunter et al., (2016) carried out restoration suitability assessment for swamps in order to safeguard and improve the provision of important ecosystem services. However, land suitability analysis of the alvars has not been performed in Estonia so far.

Therefore, the aim of this research is to determine which environmental characteristics can be used to predict suitable locations for alvar grassland habitats, and create suitability maps for potential alvar restoration regions. This work will especially be beneficial when considering the limited areal distribution of alvars in Estonia. Land suitability analysis is a frequently used technique for choosing appropriate location for an activity or for facility or to answer the question what and where it can be done (Joerin et al., 2010). For this purpose, Random Forest (RF) method of Machine Learning technique and Geographic Information Systems' (GIS) based land suitability analysis, together with Multi Criteria Decision Making approach (MCDM) was used. RF learned the given data by itself and made predictions based on the learned data. Analytical Hierarchy Process (AHP) incorporated MCDM required construction of pairwise comparison matrices, assigning importance values to the criteria and calculation weights for each criteria. Based on the calculated weights each criterion was ranked and further used in weighted overlay analysis. Performed analysis covered the whole Estonia.

The aim of the thesis was to find potentially suitable areas for restoration of habitat for alvar grassland species and related ecosystem services. These restoration areas included (1) totally new areas where alvars have never existed before but where the combination of different environmental parameters indicates that these areas can be suitable for establishment of alvar-like habitats, and (2) restoration of the areas which have been historically alvars but have been altered due to the heavy human intervention and change of land use practices. The research aim was achieved by combining different environmental variables that directly or indirectly affect alvar occurrence and persistence.

2. Theoretical background

2.1. Alvar grasslands

Alvar grasslands are biodiverse habitats where dispersed shrubs and rare tree coverage occurs. Alvars are flat, relatively open areas with shallow or sporadic soil cover (often < 20 cm) over calcareous limestone or dolomite bedrock (Albert, 2006). There is limited distribution of alvars in the world. Alvars are mostly occurring in the areas exposed to limestone bedrock. In Estonia, they are mostly found in Saaremaa, Muhu, Läänemaa, Hiiumaa, as well as in Harjumaa, Ida and Lääne-Virumaa. Two thirds of all alvars in the world occur in Sweden and one third in Estonia. They can also be found in smaller quantities in Northwest Ireland, St. Petersburg region in Russia and Great Lakes region in Canada and USA (Gazol et al., 2012). This makes alvars globally rare and emphasizes the need of their protection (Helm, Urbas, & Pärtel, 2007).

Alvar grassland environmental conditions and vegetation can vary based on their soil and moisture conditions. For instance, although all alvar grasslands are characterised by very shallow soil (less than 20 cm), some alvar habitat types have soil depth less than 5 cm or almost completely missing, exposing patches of bare rock. Under exogenic factors such as wind and/or solar radiation this shallow soil layer often dries out resulting in harsh conditions for the vegetation. On the other hand, the poor drainage characteristics of the bedrocks, where alvars are formed, results in formation of occasional and sometimes permanent water pools during rainy seasons. Frosty winter season might also affect shallow soiled alvars: open areas of soil, where snow was blown away under strong winds, start to move similar to those in arctic areas (Pärtel et al., 1999). Some alvar types might have limestone gravel and very fine upper soil layer up to few decimetres (Rosén, 1982). Alkaline soils of alvars are nutrient rich, but as they are mostly very shallow, nutrient availability for plants is low and plants have restricted growth. Because of this and very dry environmental conditions alvars are compared with steppes (Pärtel et al., 1999).

Alvars are some of the most floristically rich north-temperate habitats known (Claudia & Douglas, 1997). They host plant species from different geographic regions of the world as a result of microclimatic environmental conditions. It is likely that most current alvar grasslands are of semi-natural origin, having developed under grazing practices over thousands of years (Laasimer, 1965). Alvar communities of the natural origin can also be found but only in the areas of land uplift from sea under neotectonic land movements (Zobel & Kont 1992). Alvar vegetation mostly consists of short and stress tolerant grass layer which prefers calcareous soils. This layer is not very productive, however is very diverse (Helm, 2006).

Area of alvar grasslands has declined severely over past century, resulting not only loss of area for grassland species, but also increasing habitat fragmentation. Habitat fragmentation, habitat destruction and degradation results in three main outcomes:

- Loss of habitat
- Decrease of area of individual grassland patches
- Increasing isolation of remained patches

Generally, it can be said that if the size of habitat patches gets smaller, they can host smaller populations and colonization of these patches will decrease as well. Therefore, this process has negative impact on persistence of species diversity. Moreover, smaller populations cannot resist to increasing stochastic extinction events and they may easily become extinct. If the colonization of isolated small patches will decrease, then there will be no individuals stopping invasion of foreign species which will also result in the total extinction of the native species. Habitat loss and/or reduction of the habitat areas are increasing the necessity to stop this process and consequently manage ecological restoration process (Butaye et al., 2005).

It is important to mention that alvar communities are significant also because they hold natural and cultural heritage in European Landscapes. There are several types of grasslands; however, alvars are of utmost importance due to their landscape beauty and species richness (Westhoff, 1971). Due to extreme biological diversity and high arthropod diversity these areas are included into NATURA 2000 network as the EU Habitat Directive priority habitat type 6280*Nordic alvar and Precambrian calcareous flatrocks (Eriksson & Rosén, 2008).

Since calcareous grasslands in northern Europe are of semi-natural origin, their long term conservation requires an appropriate management. The potential for natural recovery and restoration of alvar habitats principally depends on a combination of several factors. However, in order to correctly define these factors, history and the condition of the community must be taken into account and considered as an important variable. While in sub-alpine region mown is more favoured than grazing (Butaye et al., 2005), in Estonia or Sweden clearing trees and shrubs, existence nearby seed sources, activating adequate grazing, limiting damage to the topsoil as a result of driving or ploughing are most appropriate management and restoration methods (Pärtel et al., 1999). Recent studies showed that the seed bank of alvars is very rich. The persistence of the seed bank for longer periods might make the restoration process successful (Kalamees et. al., 2012). Therefore, activation of habitat patches' network is significantly important. Individual alvar conservation sites are simply not enough taking into account urge of grassland species' genetic material exchange among different habitat sites for their long term preservation (Kuussaari, 2009).

2.2. Alvars in Estonia

In Estonia, alvars originated and developed under human influence within thousands of years, where human management, such as cutting hay or animal grazing(sheep, horses), and removal of trees and shrubs was the reason of grasslands' persistence (Pärtel et al., 1999). Alvars are species rich communities, and up to 40 vascular plant species can be found in $1m^2$ (Pärtel et al., 1999). Due to this amount of vascular plat species, in Estonia alvars are considered to be second species rich communities after wooded meadows (Kuussaari, 2006). Further, alvars in Estonia and Sweden are hosting many rare and threatened plant species (Znamenskiy, Helm, & Pärtel, 2006). Such a varied species composition of alvars determined not only by long term effects of human management of these territories but also because of the strong connectivity of alvars due to their vast distribution (Helm, 2006).

Alvars have been primarily used as pastures. Cessation of grazing in alvar grasslands since the 1950s has resulted in overgrowing of these communities (Laasimer, 1965), which lead to the gradual decrease of species richness of these special communities. Currently, because of the significant change in pattern of traditional land use practices, most of Estonia alvar grasslands are no longer used for grazing and they are overgrown with shrubs, trees and tall grazing-sensitive herbaceous plants. Even thin soiled alvars are not safe from overgrowth (Pärtel et al., 1999). Currently, ca 5000 hectares of alvar grasslands are grazed (EELIS 2019).

Alvars have experienced areal changes (Figure 1) due to different reasons. In the period of 1950 to 1980 large areas of alvar grasslands were planted with pine trees as a part of afforestation program (Kaar, 1986). In Saaremaa this program resulted in the loss of 6000 ha of alvar territories (Helm, 2006). Fertilisation and conversion to intensively managed grasslands became another enemy of alvars (Kuussaari, 2009).

Initially, in 1930s there were recorded approximately 43000 ha of alvars in Estonia. However, registers show that between 1978 and 1981 this number dropped up to 16000 ha.

Furthermore, 25 % of the remained alvars more or less are encroached by forest (Pärtel et al., 1999). It is said that once 75% of the alvars covered with shrubs, species richness will drop drastically (Znamenskiy et al., 2006). Nowadays, only ca 17000 ha of alvars are left in Estonia and ca 5000 ha are managed (EELIS).

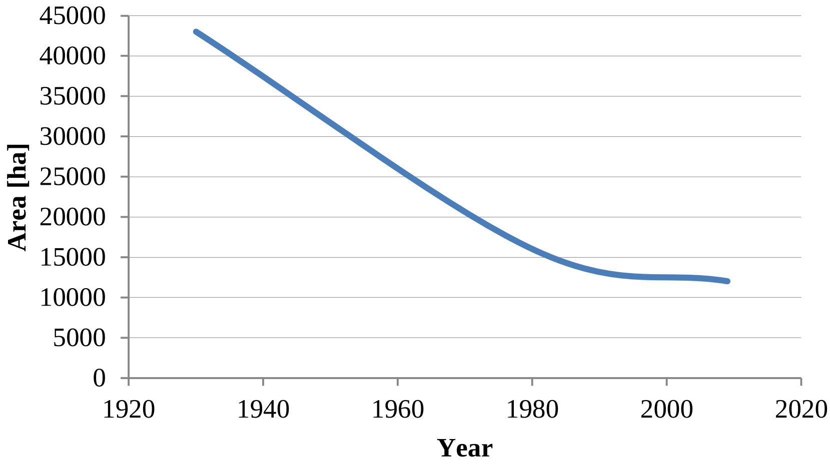


***Figure 1**. Change in the area of alvars in Estonia from 1930 to 2009*

Estonia is hosting 28% of the world's alvars. On the other hand, alvars have been confirmed as priority habitat type by Natura 2020. In Sweden alvars on the Öland Island are part of the UNESCO World Heritage sites. However, alvar protection organisation was not very active in Estonia. In recent years it became obvious that if no serious steps will be taken, it might result in total disappearance of valuable alvars (Helm, 2006). Since then ca 3000 ha of alvar grasslands have been restored, especially on the western islands of Estonia (LIFE to alvars).

Since alvars in Estonia and in Sweden are part of "traditional rural landscape", their restoration is also important from the nature conservation point of view. Furthermore, semi-natural functions of alvar communities (e.g. meadow meat production) can also bring financial profit (Rosén, 1982).

Previously alvar restoration practices were mainly based on the Swedish experience, where since the beginning of 90's approximately 7000 ha of alvars have been restored. By 2011, Estonian alvars still retained species pool while Juniper scrub coverage was already very extensive (Helm, 2006). Recently, alvar restoration measures aimed immediate clean-up of Juniper coverage in existing alvars in order to prevent further overgrowth and reduce this coverage to 30%. The biggest project involving these actions was "LIFE to alvars" that run from 2014 to 2019 involving following partners: Environmental board, University of Tartu, University of Life Sciences and the Seminatural Community Conservation Association. This action beside restoring ca 3000 ha of alvars also gained some time in order to organise proper management techniques on the alvars.

Most of the actions for alvar grasslands' restoration in Estonia involved clean-up of unwanted vegetation. However, there was yet no project or activity that considered data driven approach to the issue. This thesis focused particularly on data approach in alvar restoration.

2.3. GIS based land suitability analysis

Land use suitability analysis is one of the most frequently used techniques in environmental management. The main idea behind this method is to choose appropriate locations for an activity or for facility or to answer to the question what and where it can be done (Joerin et al., 2010). Simple premise of land use suitability analysis is that in any case there are environmental characteristics which are either suitable or unsuitable for the planned activity of the analysed situation (Parry et al., 2018). In existing literature there are many examples of its application such as agricultural land suitability (Ahmed et al., 2016), and suitability analysis of declining habitats (Busby & Whistler, 2002), environmental impact assessment (Moreno, & Seigel, 1988) and many others. Consequently, "the land suitability analysis problem involves classification of the units of observations according to their suitability for a particular activity" (Malczewski, 2004).

Land suitability has roots dating back to the late 19th and early 20th century when practitioners were using an overlay technique for hand drawn maps (Steinitz et al., 1976). As time passed and more technological innovations were done, this "overlay" technique was advanced by McHarg (1969). He suggested mapping attributes of natural and human made environment of the areas of interest and later on to represent these attributes on transparent maps by light to dark shading as high to low suitability accordingly. Afterwards, by laying every created transparent map on each other, land suitability maps of every land use were displayed. While McHarg's approach was recognized as forerunner of modern GIS overlay technique, Tomlinson's Spartan Air Services of Ottawa company was the first one suggesting computerization of overlay technique (Malczewski, 2004). Together with technological infrastructure development, overlay processes became an even more integral part of the land use suitability analysis in urban, environmental, and regional planning. Since technologies advanced, manual overlay method was replaced by computational methods and instead of storing suitable or not suitable areas in colour scale (light to dark shading) they started to store results in numerical values as matrices in computers (Murray et al., 1971). Boolean operations and/or weighed linear combination were the most frequently used methods for suitability analysis since they were easy to understand and implement in available GIS software although these methods were heavily criticized due to issues of independence among criteria chosen for suitability and standardization of suitability maps (Malczewski, 2004). It was also considered by scholars that Boolean operations and weighed linear combination is simplifying complicated suitability analysis processes since they focus just on certain facts instead of focusing on combination of facts and value judgements (Malczewski, 1999).

It is said that the quality of the planning process directly depends on the availability of the data and existence of proper and reliable data processing tools. The better data processing is the better planning results will be because "planning is fundamentally a sequence of rational and technical procedures" (Hall, 1974). Since the nature of the planning process has changed from just being a scientific approach and now involving to the decision making process also non-experts in this field, such as stakeholders, communities, interest groups and others, it increased the role of GIS in the planning process. Actually, changes in the planning processes were paralleled with the changes, better accessibility, of GIS technologies.

Technological advances affect all the main components of GIS tools such as data input, data storage, data analysis and spatial data output. Nowadays, there is also a vast amount of GIS software which is available to use on any kind of computers and they are improving very fast as the advancement in information technologies goes on (Malczewski, 2004). Therefore, GIS is distinguished from other systems because of its capabilities to execute combined analyses of spatial data and attribute data and therefore to develop alternative scenarios (Parry et al., 2018). GIS has capacity to integrate different data (soil, climatological, hydrological and etc.) which later can be used for obtaining information for different application purposes (land use suitability and etc.) by manipulating and analysing input data (Puntsag et al., 2014). Such capability of GIS systems makes them easier for the users to deal with data and to make reliable conclusions in a considerably easier way.

Depending on the GIS systems and the purpose of the analysis, especially in the case of land use suitability analysis, it is better to differentiate between two categories of GIS operations: fundamental and advanced operations. Fundamental or basic operations are the ones that can be done in most GIS software and include overlay and scalar operations, measurements, connectivity and neighbourhood operations. However, in order to be useful for decision making process, GIS should also provide range of advanced or compound operations. For instance, cartographic (spatial) modelling can be counted as one of advanced GIS operations (Malczewski, 2004). Cartographic modelling is basis of the land use suitability analysis that was developed to plan land use alternatives by analysing several geographically distributed factors (Tomlin, 1990). Cartographic modelling method is organizing fundamental operations of GIS into complex spatial models. Additionally, many GIS provide programming languages (script) while others provide graphical environment (flowchart approach) for executing spatial operations and cartographic modelling. Lastly, capability of GIS is to support decision making processes which makes it of particular importance for land use suitability analysis and modelling (Malczewski, 2004).

2.4. Land suitability analysis as a tool for restoration and conservation activities

Due to the current rates of habitat fragmentation, degradation and loss, many species face severe risk of extinction. Initially, most of the existing literature addressed issues of spatial pattern and arrangement in terms of species persistence. However, not many of them included how spatial pattern can be used for instance for species recovery plans. Huxel & Hastings (1999) have suggested to include spatial processes into restoration management plans to reduce the effects of habitat loss and fragmentation. Another important question, besides which habitat should be restored, is how much of the habitat has to be restored (Huxel & Hastings, 1999).

With increasing urge of habitat conservation, growing number of literature also became available on land suitability analysis of different land uses for protection and conservation purposes due to the environmental services they provide. Some of these studies focus on mapping the distribution of species in relation to protected areas. Others focus on gap analysis, by identifying gaps and thus threats for long-term conservation of certain habitats or species in particular. There are also studies especially focusing on the suitability analysis of habitats which's species and communities are of high importance for restoration and conservation (Geneletti & Duren, 2008).

Novak and Short (2000) performed suitability analysis of Eelgrass meadows in Plum Island. The main reason of this analysis was to prepare restoration basis of eelgrass habitat since there was recorded significant decline in species composition of the meadows. Different attempts were made to restore eelgrass, but mixed results were achieved largely because of the poor site selection. Therefore, site selection was considered to be the most important factor in successful eelgrass restoration (Fonseca, 1998). Because of the catastrophe of Berman Oil spill over reefs, habitat suitability analysis was performed as a part of restoration planning since due to the discharged oil, in total 1,009 m^2 area of reefs near San Juan, Puerto Rico had been affected and the eolianite reef was scarified (Jack & Suite, 2005).

Considering, nearly all of the species are suffering from one or many changes happening in their natural habitats, Fernandez and Morales San Martin (2016) performed land suitability analysis and found potential areas for restoration of threatened endemic species (*Bielschmiedia miersii* and *Pouteria splenden*). This study also took into consideration future climate change. As swamps provide significant ecosystem services, such as improving water quality, storing flood water, providing habitat for wildlife, storing carbon, their restoration suitability assessment was carried out by Hunter et al., (2016) in Pontchartian Basin, Louisiana. Consequently, swamp areas, areas ready for restoration, areas where swamps can be potentially restored in the future and areas where restoration is not recommended were detected. Ouyang, Lu et. al., (2011) and Uuemaa et. al., (2018) did GIS based suitability analysis for wetland creation and restoration in Yongding River, China and South Island, New Zealand respectively. As wetlands are providing valuable environmental services, creation of new wetlands and/or restoration of existing wetlands were subjects undergoing intense study. However, these studies included only measures such as improving water quality or protecting biodiversity of these habitats. In this paper authors are emphasizing importance of evaluation of wetlands on the watershed level for finding suitable sites for their restoration. As a result, areas with highest and lowest suitability for wetland restoration were identified.

2.5. Methodology of suitability analysis

2.5.1. Multi-criteria decision making

In many studies incorporation of the widely known method MCDM into land suitability analysis helped to reduce the oversimplification problem of this analysis (Pohekar & Ramachandran, 2004). MCDM is a method that supports decision makers in combining several options, where the main concern is a combination of information from several criteria to form a single index of evaluation (Ahmed et al., 2016). The process of the MCDM approach consists of several interdependent steps that result in the final decision on the studied topic (Pohekar & Ramachandran, 2004). In other words, it helps to set the relationship between input and output data. If MCDM problem has M alternatives and N criteria, then we can express it in a matrix shown on the Figure 2.

$$\begin{array}{cc} & \begin{array}{cccc} w_1 & w_2 & \cdots & w_n \\ C_1 & C_2 & \cdots & C_n \end{array} \\ M = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{pmatrix} & \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{pmatrix} \end{array}$$

Figure 2. *Matrix of MCDM problem where A_1, A_2, A_m are relevant alternatives, C_1, C_2, C_m are the criteria, Z_{ij} is the performance value of alternative A_i under criterion C_j and w_j is the weight of criterion C_j.*

This method is also advancing the traditionally known overlay method behind the land use suitability analysis (Malczewski, 1999). Furthermore, the combination of GIS-MCDM is a widely used technique since GIS is capable of handling a wide range of criteria from various sources in time and cost- efficient analysis (Chen et al., 2010).

The most frequently chosen method from various MCDM methods is analytical hierarchy process (AHP). AHP can help to defining the weights of each criterion involved in the process. This technique introduced by Thomas Saaty (1986) is one of the most accurate approaches to calculate weights of criteria. It is a well-known and useful approach in cases where many alternatives/criteria are available. AHP can manage different criteria into a hierarchy tree where the upper level is the problem to which a solution is looked for and the lowest level contains various sub criteria or parameters. It is a square matrix based on a pair-wise comparison procedure of the criteria, where the number of rows and columns is defined by the number of criteria to weight. Based on an expert opinion each criteria receives relative importance value following the fundamental scale of absolute importance numbers by Saaty (2008, Table 1). The consistency ratio and weights of each criterion are derived from the assigned importance values. This method has also been incorporated into GIS based suitability procedure (Chandio et al., 2013). Therefore, it makes the execution of land suitability analysis in GIS, using weights obtained from AHP, relatively easy.

Table 1. *The fundamental scale of absolute numbers by Thomas Saaty (2008)*

Intensity of Importance	Definition	Explanation
1	Equal Importance	Two activities contribute equally to the objective
3	Moderate importance	Experience and judgement slightly favour one activity over another
5	Strong importance	Experience and judgement strongly favour one activity over another
7	Very strong or demonstrated importance	An activity is favoured very strongly over another; its dominance demonstrated in practice
9	Extreme importance	The evidence favouring one activity over another is of the highest possible order of affirmation
Reciprocals of above	If activity i has one of the above non-zero numbers assigned to it when compared with activity j, then j has the reciprocal value when compared with i	A reasonable assumption
1.1–1.9	If the activities are very close	May be difficult to assign the best value but when compared with other contrasting activities the size of the small numbers would not be too noticeable, yet they can still indicate the relative importance of the activities.

By summarizing and concluding reviewed literature, following steps in land suitability workflow were established:

- Choosing important criteria for alvar suitability analysis
- Assigning importance to each criteria and calculation of weights
- Using calculated weights in Weighted overlay analysis
- Finding the most suitable areas for alvar restoration

2.5.2. Random forest

Machine learning methods become more and more popular in land suitability analysis due to their capability to deal with complex relationships between predictor variables, robustness in managing big and noisy data and being not very time consuming (Lahssini et. al., 2015). Several machine learning techniques have already been incorporated into land suitability analysis. For instance, Wen et. al. (2009) used classification and regression tree to investigate hydrological requirements of the river Red while Park et. al. (2003) applied artificial neural network to predict aquatic insect species. Another try has been given to study landscape configuration and habitat suitability using genetic and simulated annealing algorithms by Holzkaemper et. al. (2006). However, many studies showed that RF (hereafter RF) has higher frequency of reaching the best predictive performance (Benito et. al., 2006). Lahssini et. al. (2015) and Vincenezi et. al. (2011) used RF for cork oak suitability and *Ruditapes philippinarum's* potential spatial distribution assessment respectively. Probability of correct predictions in both studies was higher than 90%. Consequently, RF was also chosen to use in this study.

RF, as proposed by Breiman (2001), "is a classifier consisting of a collection of tree - structured classifiers {h(x,Θk), k=1, ...} where the {Θk} are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x". This method is considered as an extension of classification and regression trees and it uses the Classification and Regression tree algorithm (CART). The Classification algorithm predicts continuous values in the form of probability for a class label (0/1) whereas the regression algorithm predicts a discrete value in the form of integer quantity (Strecht et. al., 2015). RF is a decision tree based classifier and can be described as trees where branches formed by the answers to yes/no questions and are not pruned (but can be). Each tree in the forest constructed using bootstrap samples from the original dataset. It uses random selection of explanatory variables or factors to split the tree at nodes, instead of splitting each node based on the best split among all the variables, thus avoiding overfitting. The goal of RF is to identify the best model to analyse the relationship between dependent and independent variables (Friedman et al. 2003). RF is proven to be also a suitable classification method when there is a correlation between the variables used for classification (Georgian et. al., 2019).

RF can be executed using different programming languages. In this thesis, scikit-learn library of Python was used (Cournapeau, 2007). According to the capabilities of the library, different parameters in the model can be tuned in order to achieve the best possible model while the usage of the default parameters might also lead to acceptable results. Some parameters that can be tuned are the following: (1) the number of trees that will make up the whole RF (n_estimators), (2) maximum depth of each tree, meaning how much each tree will expand (max_depth), and (3) minimum number of samples required to split an internal leaf node to have a more specific classification (min_samples_split).

RF, both in classification and regression models, also provides a measure of the variable importance based on the contribution of the variable to the model at each node and each tree where it appeared. Another estimate value that can be obtained from the model is the “out of bag” (OOB) score, an average error of prediction of out of bag samples (samples that do not appear in bootstrap samples, Breiman, 2001).

In order to evaluate the goodness of the RF model, the data needs to be split into two parts: training and testing data. This helps to evaluate the performance of the algorithm for the chosen problem by training one sample of data and validating it on the test sample. The proportion of train/test dataset needs to be chosen wisely in order to avoid overfitting. Meaning the model can learn not just the actual relationship in the training data but also the noises present in the data. Furthermore, it can memorize the data (Breiman, 2001).

3. Data and methods

The majority of the studied alvars are located in the Western and Northern parts of Estonia (Figure 3). Considering such restricted areal coverage and their high importance, it is necessary to study whether it is possible to restore them also outside of their common occurrence area. Therefore, the study area covers the whole inland area of Estonia, summing up to approximately 43000 km^2.

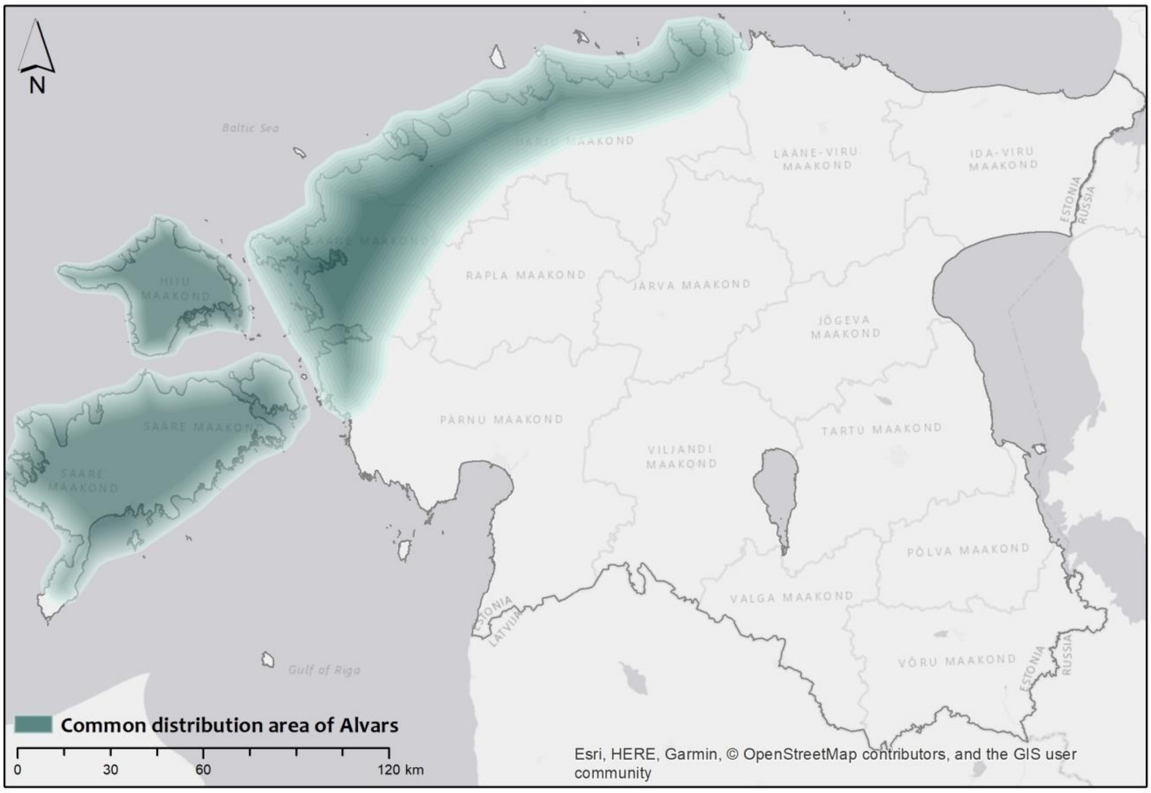


Figure 3. *General occurrence area of alvars in Estonia based on the existing data*

3.1. Data

3.1.1. Alvar distribution data in Estonia

Alvars are not explicitly distinguished on the available land use maps and are rather included into the class of grasslands. Therefore, botany department of UT provided the data for the land suitability analysis. Two major datasets in a form of polygon layers were made available. One of them contained quite current information on alvar distribution. This dataset is a result of the survey of the Estonian Semi Natural Community Conservation (2000-2010) and alvar distribution mapping based on the Estonian state-run database EELIS. Another dataset resulted from the Estonian vegetation mapping from 1930 to 1950. Consequently, the first dataset was used to understand the current situation in alvar distribution. While the latter one was helpful in understanding historical distribution of alvars.

Attribute table of the historical distribution layer was comprised of field ID (FID), x and y coordinates of polygons and the area values calculated in hectares. Attribute table of the current alvar distribution layer contained the same information as the historical layer as well as the county alvar is located on and the conservation category information (A, B and C).

3.1.2. Environmental variables

Areas suitable for restoration of alvars in whole Estonia were searched based on the combination of different environmental variables that are shown in Table 2. The environmental predictor variables were extracted from Estonian Soil Database, LiDAR based digital elevation model (DEM) and Estonian Digital Topographic Database (ETAK). All these datasets are open data and can be downloaded from Estonian Land Board (Estonian Land Board). For the final analysis, updated and improved Estonian Soil Database by Kmoch et al. (2019) was used. From the soil database, the soil types and soil textures that most frequently occur under alvars in the past and present were extracted. The DEM was used to calculate slope and a Topographic Wetness Index (TWI). TWI is a well-known terrain derivative in ecology and hydrology. It shows a relative measure of moisture status in a given area or pixel (Buchanan et. al., 2013).

The Digital Topographic Database of Estonia was utilized to extract the land use information. Originally, this database consists of separate layers comprised of point, line and polygon objects. Due to the purpose of this work, only the land use information in the form of polygon objects were used.

Table 2. *Environmental datasets and variables used in land use suitability predictions*

Datasets	Source	Predictor variables
Soil database	Estonian Land Board & Kmoch et. al., (2019)	Soil type Soil texture Soil depth
DEM	Estonian Land Board	Slope TWI (calculated from DEM)
Digital Topographic Database	Estonian Land Board	Land use

To prepare the data for further use, there were several general hypotheses posed:

- Since alvars are calcareous grasslands on very shallow soils, soil depth is expected to be the most important variable for alvar identification.
- It can also be expected that there are certain soil types (e.g. Rendzik Leptosols, Calcaric Regosols, Calcari Abruptic Gleysols etc.) and soil textures (e.g. clay and/or sandy) associated with alvars in the past and nowadays.
- Alvars can be found only on outcrops or three bedrocks, Ordovician, Cambrian and Silurian. Thus, this information will definitely contribute to the all upcoming analysis by restricting suitability search area.

3.2. Methodology

3.2.1. MCDM for alvars' suitability assessment

Overall flowchart of MCDM based alvar suitability analysis is shown in Figure 4. The existing data was examined in terms of criteria extraction. Six criteria, including soil type, soil texture, bedrocks, slope, TWI and land use were used in land suitability analysis, as described in the following.

Soil type

In order to do pairwise comparison of different soil types, the most preferred soil types by alvars were identified. In total, eight frequently occurring soil types were identified. These types were used both in pairwise comparison matrices as well as in reclassification of GIS layers.

Soil texture

Similarly, to soil type, four main soil textures from the existing ones were drown out for further pairwise comparison due to their frequent occurrence under alvars.

Bedrock

Reviewed literature pointed out that alvars are characterized with the three main bedrocks: (1) Silurian, (2) Cambrian and (3) Ordovician. They were given priority during the pairwise comparison process.

Land use

In terms of suitability of different land uses, expert opinion was taken into account. Expert judgments were based on the particular preferences of alvars as suitable habitat as well as the general existing situation of land use in Estonia.

According to the acquired information and on expert opinion, pairwise comparison matrices for all the criteria were formed. Importance of each criteria over another criteria was estimated following the fundamental scale of absolute importance numbers by Saaty (2008), which is shown in Table 1. Similar steps were repeated for each criterion.

An important part of the pairwise comparison procedure is calculation of consistency ratio (CR) which helps to minimize bias in criteria weighting. This ratio shows "how consistent judgements have been relative to large samples of purely random judgements" (Saaty, 2008). According to Saaty (2008), only a CR below 0.10 (10%) can be considered as acceptable and can guarantee that the subjective judgement of the expert was correct. If the CR was higher than 0.10, expert matrices were revised until the desired value has been obtained. CR can be calculated in two steps shown in the equations below:

$$1)\ \mathrm{C.I} = \frac{\lambda_{\max} - n}{n-1} \qquad\qquad 2)\ \mathrm{C.R} = \frac{\mathrm{C.I}}{\mathrm{R.I}}$$

where **C.I** is the consistency index, **n** is the number of items being compared in the matrix, **max λ** is the largest Eigen value and **RI** is random consistency index.

Pairwise comparison was executed in MS Excel. Final weights of each criterion are shown in Table 3, while the weights of the criteria are in Table 4.

Table 3. *Main criteria used in alvar suitability analysis with their weights*

Criteria	Weights
Soil type	*0.194*
Soil texture	*0.030*
Bedrock	*0.485*
Slope	*0.080*
TWI	*0.032*
Land use	*0.179*

Table 4. *Criteria with their criterion and with the calculated weights*

Criteria	Criterion	Weights
Soil type	K	0.274
	Kr	0.205
	Kh	0.158
	Gh	0.117
	Khg	0.090
	Kg	0.066
	Gk	0.051
	Go	0.038
Soil texture	SL	0.520
	LS	0.268
	S	0.141
	L	0.071
Bedrock	Silurian	0.171
	Cambrian	0.422
	Ordovician	0.269
	Vend	0.069
	Devon	0.069
Land use	Scrubland	0.208
	Cropland	0.052
	Forest	0.244
	Grassland	0.386
	Wetland	0.023
	Urban	0.086

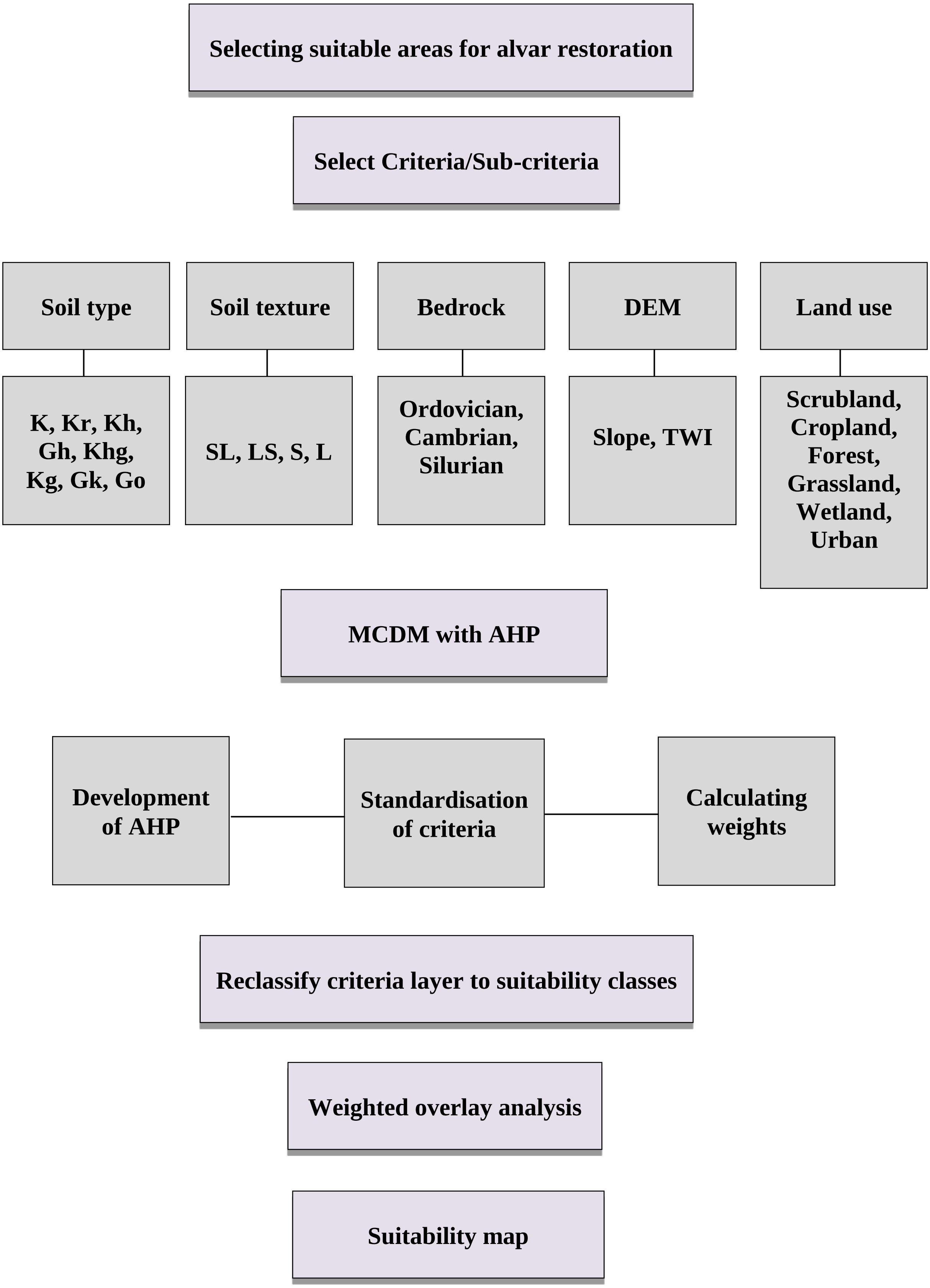


Figure 4. Flowchart of Methods with MCDM

3.2.1.1. Weighted overlay analysis

Weighted overlay analysis is a technique that allows users to apply common scale of values to dissimilar data in order to achieve integrated analysis.

After obtaining weights for each criterion, attributes of all the GIS layers were reclassified taking into account criterion weights. Afterwards, these layers were converted into raster layers with a resolution of 30 meters. Weights of each criterion were used in final weighted overlay analysis in order to prioritize one criterion over another one. All the raster layers were overlaid using Weighted overlay tool in ArcMap 10.6. This tool reclassifies values in the input raster into common scale. Afterwards, it multiplies the values in the cells of raster with the importance weight of the same raster and combines the cells together where the calculated value is the same. However, the tool accepts only integer rasters.

In this study, the following four suitability classes were differentiated:

- **Highly suitable** - lands having no significant limitation for alvar restoration (“4”)
- **Moderately suitable**- lands having some limitations for alvar restorations (“3”)
- **Marginally suitable**- lands with extreme limitations for alvar restoration (“2”)
- **Unsuitable** - lands not suitable for alvar restoration (“1”)

3.2.2. Random forest model for alvars’ suitability assessment

Prior to building RF models, the normality of the variable distributions was checked. Histograms and Shapiro test showed that most variables were not normally distributed. For preliminary detection of simple relationships between different variables, Spearman’s rank correlation was computed for all the variables. Further, RF technique was applied in the order showed on Figure 5.

For finding suitable areas using RF, a list of variables (predictors) is given in Table 5. As a first step, categorical variables were converted into dummy variables.

Table 5. *List of variables involved in RF models*

Predictors	Data type
Soil type	Categorical
Soil texture	Categorical
Soil clay content	Numerical
Soil silt content	Numerical
Soil sand content	Numerical
Soil rock content	Numerical
Slope	Numerical
TWI	Numerical
Bedrock	Categorical

In order to be able to decide which variables from Table 5 are most suitable for alvar suitability assessment, five different RF models were constructed by combining different sets of predictors. The general procedure in RF is to split the data into two parts: training/test.

Using the training set, model is trained to recognize required patters, in this case alvars. In order to assess the performance of the trained model (how well it can recognize alvars), test sets are used. Since the aim of this work is to find suitable areas throughout Estonia, datasets containing alvar locations were used for training/test purposes. The dataset covering the rest of Estonia was used to make final suitability predictions. According to existing literature, the proportion of training/test is crucial, because models can learn too much (or not enough) and the prediction performance will not be realistic.

Consequently, five models (Table 6) with two different training/test split proportions (60:40 and 70:30) were tested. Using "RandomizedSearcCV" function from scikit-learn library (Cournapeau, 2007), the best set of parameters(n_estimators, max_depth etc.) for the models was searched. The target variable in the training phase was the alvar data. At the end of this step, performance of each model was examined and the best model was selected. OOB (out of bag) score is a way of validating RF models and it is a measure of how successful prediction was. At this stage it is also possible to check influence of each predictor variable to the model's information gain.

Lastly, using parameters of the best model and the dataset covering whole Estonia, suitable areas for alvar restoration were identified. This was a regression task, meaning when probability of each polygon being alvar was higher than 0.9, it was accepted as a highly suitable polygon(area) for alvar restoration.

RF models were built and executed using scikit-learn library in Python and the results of the RF were mapped using ArcMap 10.6.

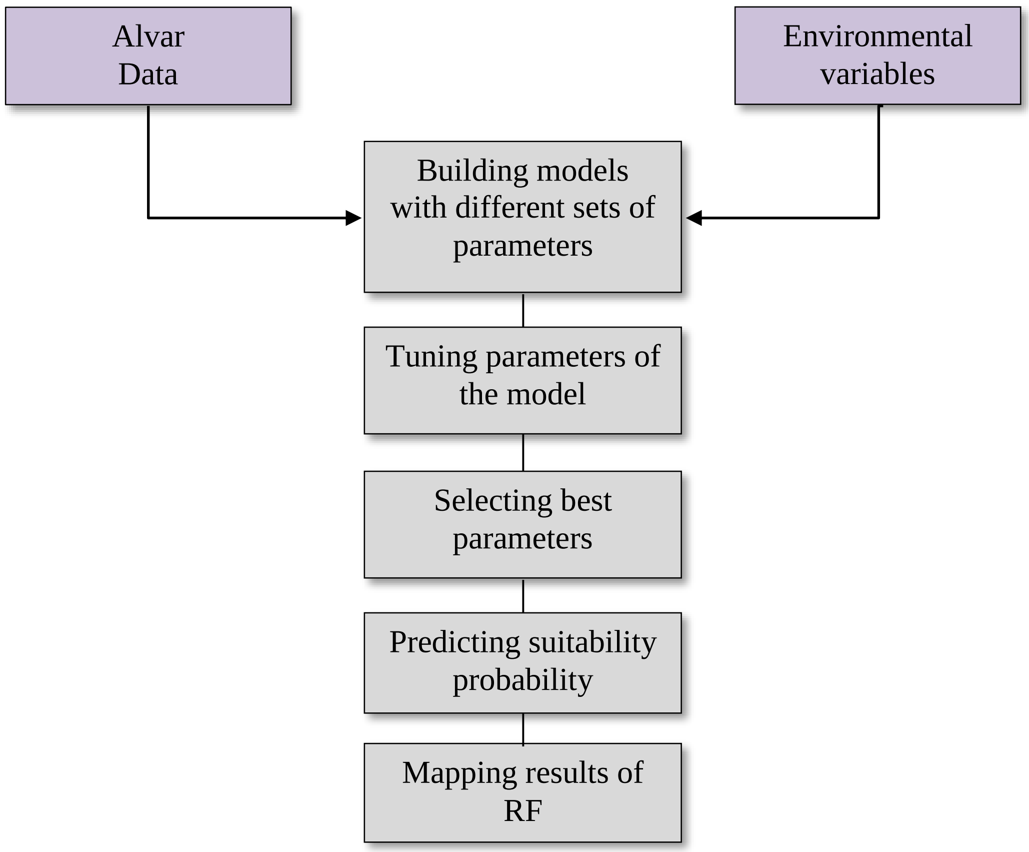


Figure 5. *Flowchart of RF model for alvar suitability assessment*

Table 6. *Table of the base set of variables used in train/test models, as well as split options and the chosen set of variables with the best split option*

Base set of predictor variables used in train/test models	Split options
[Model 1] Soil type, Soil texture, Soil clay, Soil silt, Soil sand, Soil rock, Slope, Topographic Wetness index, Bedrock	**[Split 1]** 60/ 40 & **[Split 2]** 70/ 30
[Model 2] Soil type, Soil clay, Soil silt, Soil sand, Soil rock, Slope, Topographic Wetness index, Bedrock	
[Model 3] Soil type, Soil texture, Slope, Topographic Wetness index, Bedrock	
[Model 4] Soil texture, Soil clay, Soil silt, Soil sand, Soil rock, Slope, Topographic Wetness index, Bedrock	
[Model 5] Soil type, Soil texture, Soil clay, Soil silt, Soil sand, Soil rock, Slope, Topographic Wetness index	

4. Results

When comparing historical coverage of alvars with the nowadays, it was revealed that the majority of alvars that had existed in 1930 to 1950ties had disappeared by 2016 ("lost alvars) (Figure 6a and 6b). However, it was also noticed that there are some areas, even though in the form of small patches, that were not shown as alvars in1930-1950 (simply unmapped or because of a different habitat classification historically and nowadays or because they did not exist) but were mapped as alvars in 2016 (Figure 6c, 6d). In this thesis, these alvars were called as "historically unmapped and/or not existing" alvars.

Preliminary statistics on the alvar distribution datasets showed that historically alvars mostly occurred on Rendzik Leptosols (Estonian abbreviation: Kh), Rendzik Leptosols + Calcaric Regosols(K), Skeletic Leptosols (Kr), Calcari Abruptic Gleysols (Gh) and Rendzi- Gleyic Leptosols (Khg). Whereas existing nowadays alvars are distributed mostly on the Leptosols+calcaric Regosols (K), Skeletic Leptosols (Kr), Rendzi-Lithic Leptosols (Kh), Rendzi-Gleyic Leptosols + Calcari Gleyic Regosols(Kg), Mollic Gleysols (Go) (Table 7).

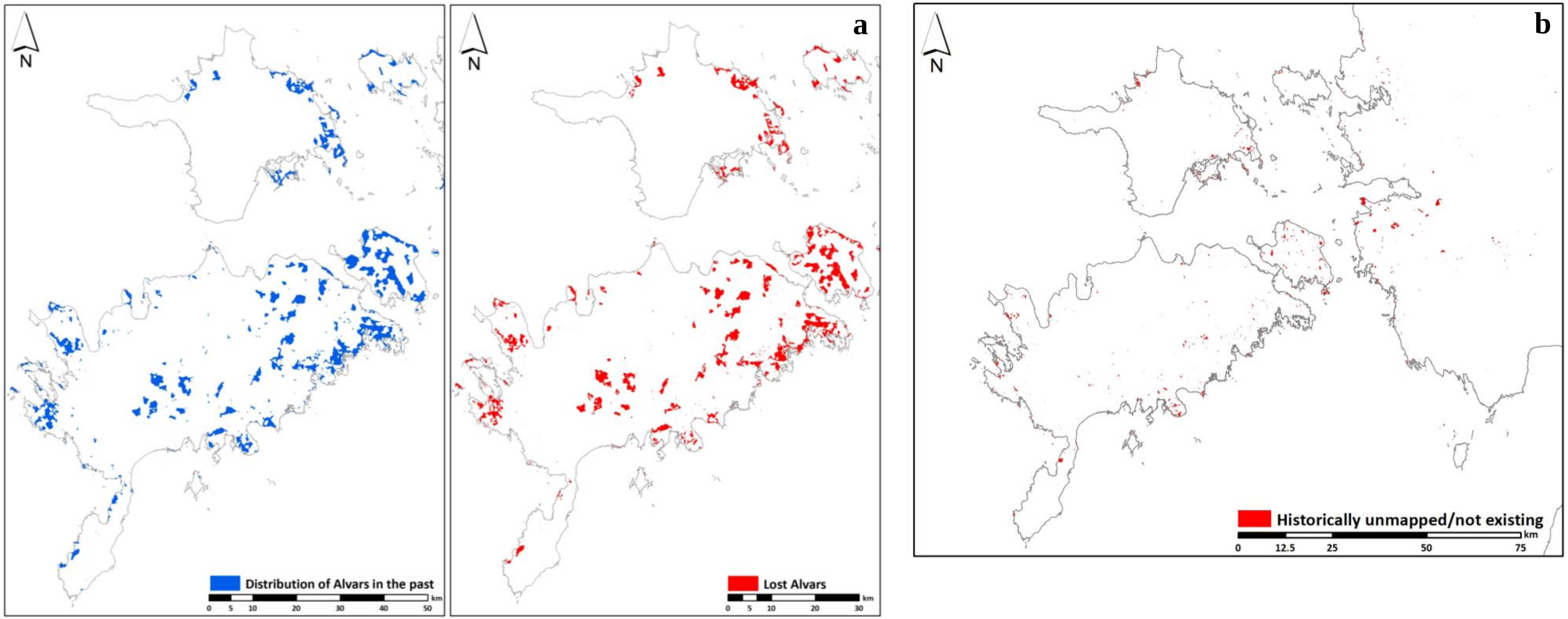
Table 7. *Area and areal percentage of the most common soil types under alvars in the past and nowadays*

Alvars	Soil type	Areal percentage	Area (m^2)
Historical alvars	Kh	14.85	2159
	K	12.89	1874
	Kr	12.54	1824
	Gh	10.42	1515
	Khg	9.71	1412
Nowadays existing alvars	K	12.52	5886
	Kr	9.46	4447
	Kh	8.76	4118
	Kg	7.83	3682
	Go	7.74	3640

Historically alvars predominantly existed on clay sand (Estonian abbreviation: sl) and sandy clay textures (ls). However, it was observed that currently existing alvars are located mostly on the clay sand (sl) and clay (s) textures. For both historical and nowadays existing alvars sandy (l) texture was also a relevant texture (Table 8).

Table 8. *Area and areal percentage of the most common soil textures under alvars in the past and nowadays*

Alvars	Soil texture	Areal percentage	Area (m^2)
Historical alvars	sl	42.05	6115
	ls	21.80	3170
	l	17.39	2529
	s	13.71	1993
Nowadays existing alvars	sl	41.79	19657
	s	18.52	8710
	l	15.39	7238
	ls	15.37	7228



N
Distribution of Alvars in the past
0 5 10 20 30 40 50 km
N
a
Lost Alvars
0 5 10 20 30 km
N
b
Historically unmapped/not existing
0 12.5 25 50 75 km

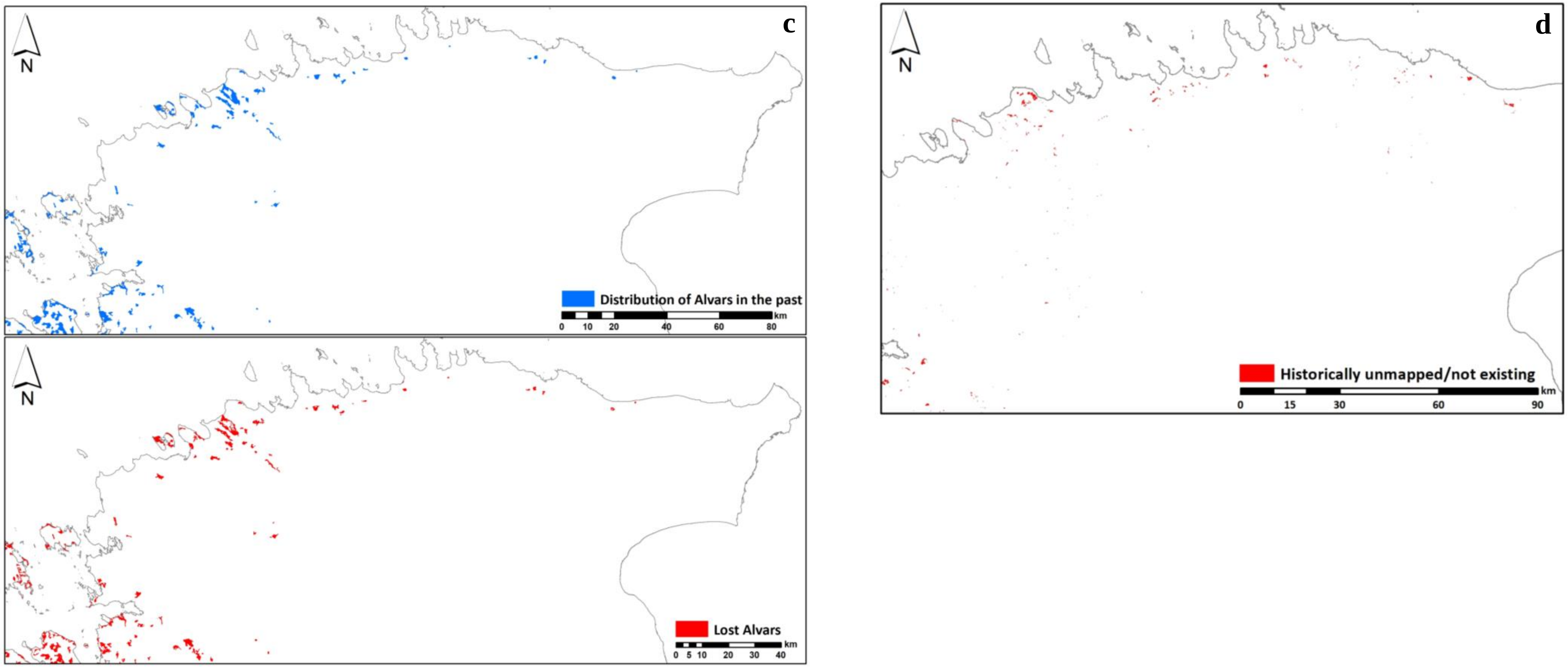


Figure 6. *Map of the distribution of the alvars in the past and lost alvars on Estonian islands (a), in the northern Estonia (c), as well as historically unmapped alvar patches on the Estonian islands (b) and Northern Estonia (d).*

From the available alvar distribution datasets, both in the past and nowadays, it was observed that alvars were located on the Silurian, Cambrian and Ordovician bedrocks.

Majority of alvars is located on slopes of 0-1.5 degrees. Therefore, these slopes were considered as the most optimal for alvar occurrence. Further, topographic wetness index of 8-11 were the most suitable whereas index between 0-8 and 11- 23 were mostly unsuitable for alvar occurrence.

Most historical alvars have become forested (48%). Furthermore, 11% of alvar areas are occupied by agricultural fields and 10% by shrublands. Only 15% of the area remained alvar grasslands.

4.1. MCDM

MCDM approach was applied together with Analytical Hierarchy process in alvar suitability analysis. There are always environmental characteristics which are either suitable or unsuitable for the planned activity or analysed situation (Parry, 2018). Based on this, most relevant environmental variables were derived from the existing datasets and their importance level was assessed through the AHP (Saaty, 1986). Calculating weights for the main six criteria helped to prioritize one criterion over another one in weighted overlay analysis. Further, weights of each criterion were used for reclassification of GIS layers in weighted overlay analysis. As Table 3 indicates, bedrocks, soil type and land use were the most important factors affecting alvar suitability analysis. This list was followed by slope and soil texture with the topographic wetness index sharing almost equal importance. The whole area was classified into four suitability classes with the class “4” being highly suitable areas and class “1” being not suitable areas. Finally, the results have been plotted and visualized.

MCDM approach predicted suitable areas in almost each part of Estonia, while the most suitable areas were found in in western islands of Estonia (Saaremaa, Muhu, Hiiumaa) as well as in north-west inland areas and northern Estonia. Majority of low/not suitable areas are located in the Southern Estonia.

Although different land uses were prioritized during the weighted overlay analysis, additional examination of the results was carried out. Consequently, 60% of the areas, classified as highly suitable, are forested areas (Table 9). Furthermore, 29% of the highly suitable areas for alvar restoration are actually grasslands nowadays. This list is followed by 4% and 3% for shrublands and croplands respectively. Very few predictions were made also on urban and wetland areas.

Table 9. Current land use in the predicted suitable areas for alvar restoration with MCDM

Landuse	Areal percentage (%)
Forest	60.25
Grassland	28.46
Shrubland	4.06
Cropland	3.44
Other	2.50
Urban	1.11
Wetland	0.13
Water	0.05

Eventually, MCDM approach in land suitability analysis of alvars predicted 987.93 km^2 of highly suitable areas for alvar restoration not including nowadays existing alvars. 207 km^2 of the predicted areas were once occupied by alvars in the past but have been altered due to the land use change.

For accuracy assessment first the result of all four classes was utilised. 98.7% of nowadays existing alvars were correctly identified by MCDM (Figure 7). However, if we only take into account the “highly suitable”-class, just 27.4% of nowadays existing alvar areas are identified (Figure 8).

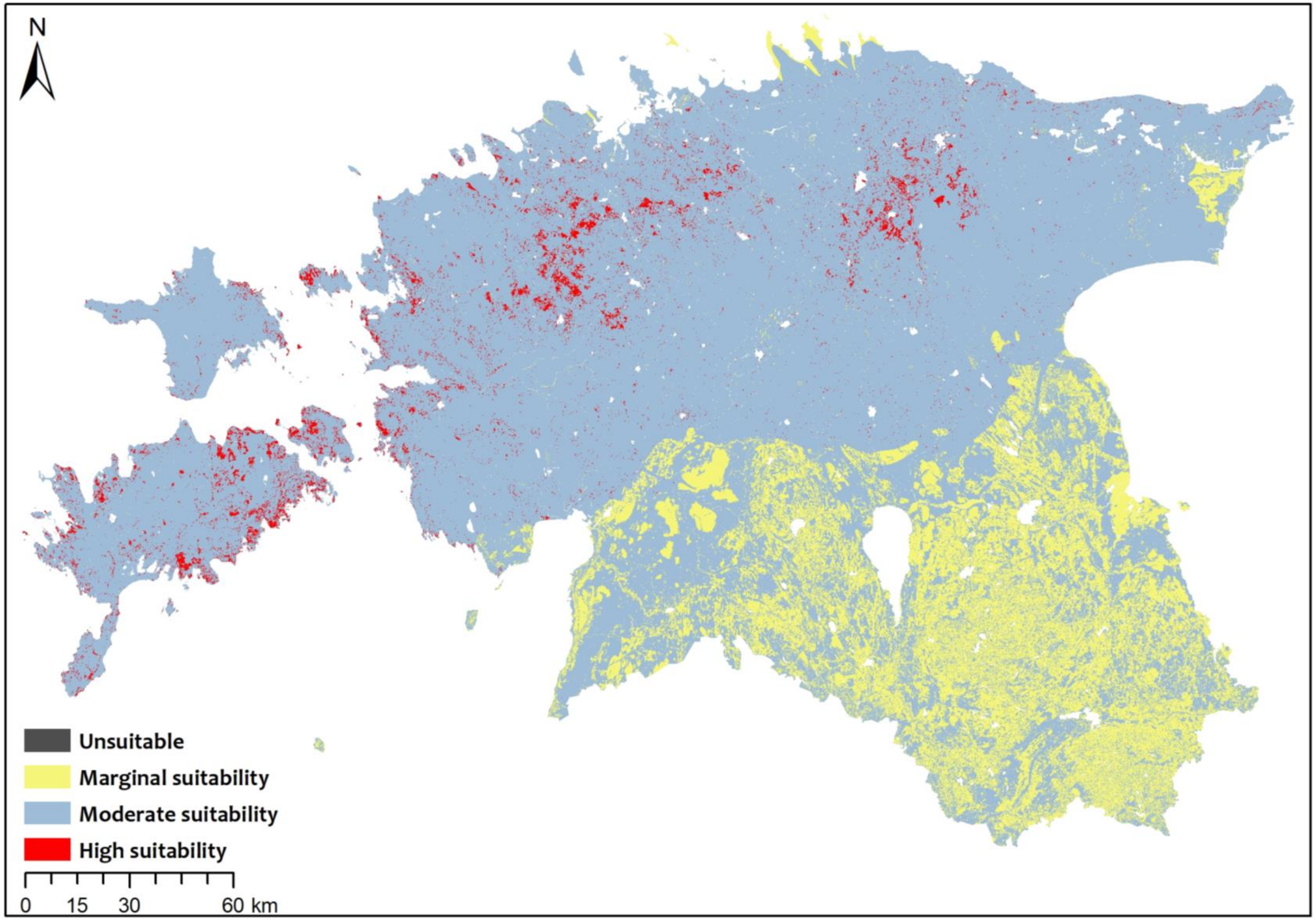


Figure 7. *Results of four classes suitability assessment with MCDM*

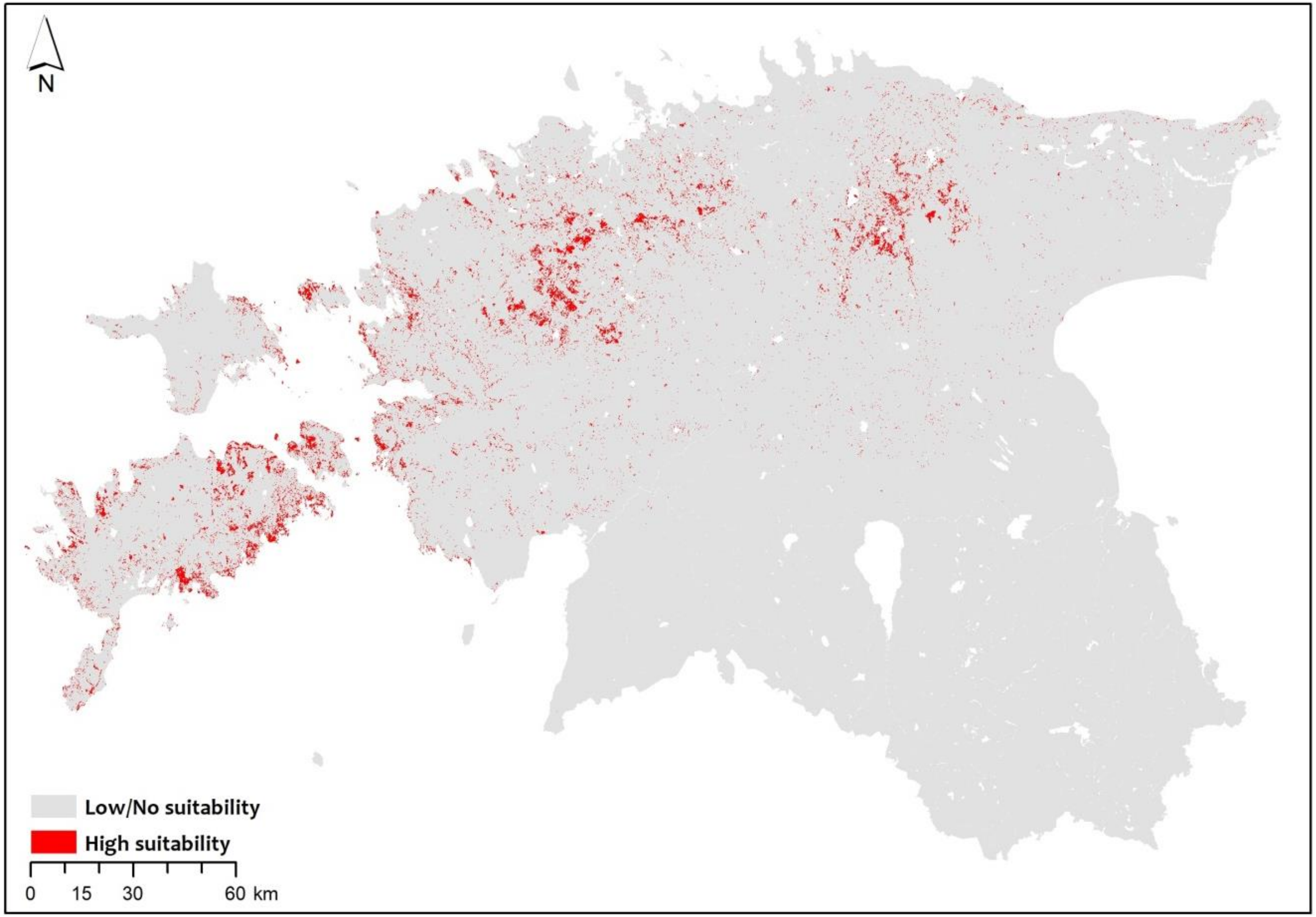


Figure 8. *Merged **r**esults of MCDM approach in land suitability analysis of alvar grasslands in Estonia*

4.2. Random forest

Prior to execution of RF models, correlation between predictor variables was measured. Several variables were statistically significantly correlated, either positively or negatively (Figure 9).

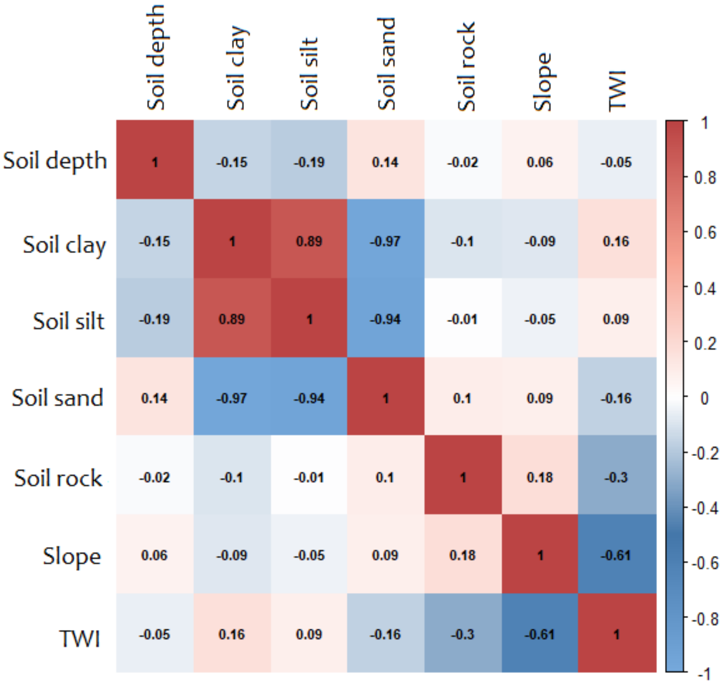


Figure 9. *Correlation matrix of numerical predictor variables*

Considering RF is capable of handling correlated variables, RF algorithms were run on the pre-selected models. The decision making process in the RF is depicted in Figure 10.

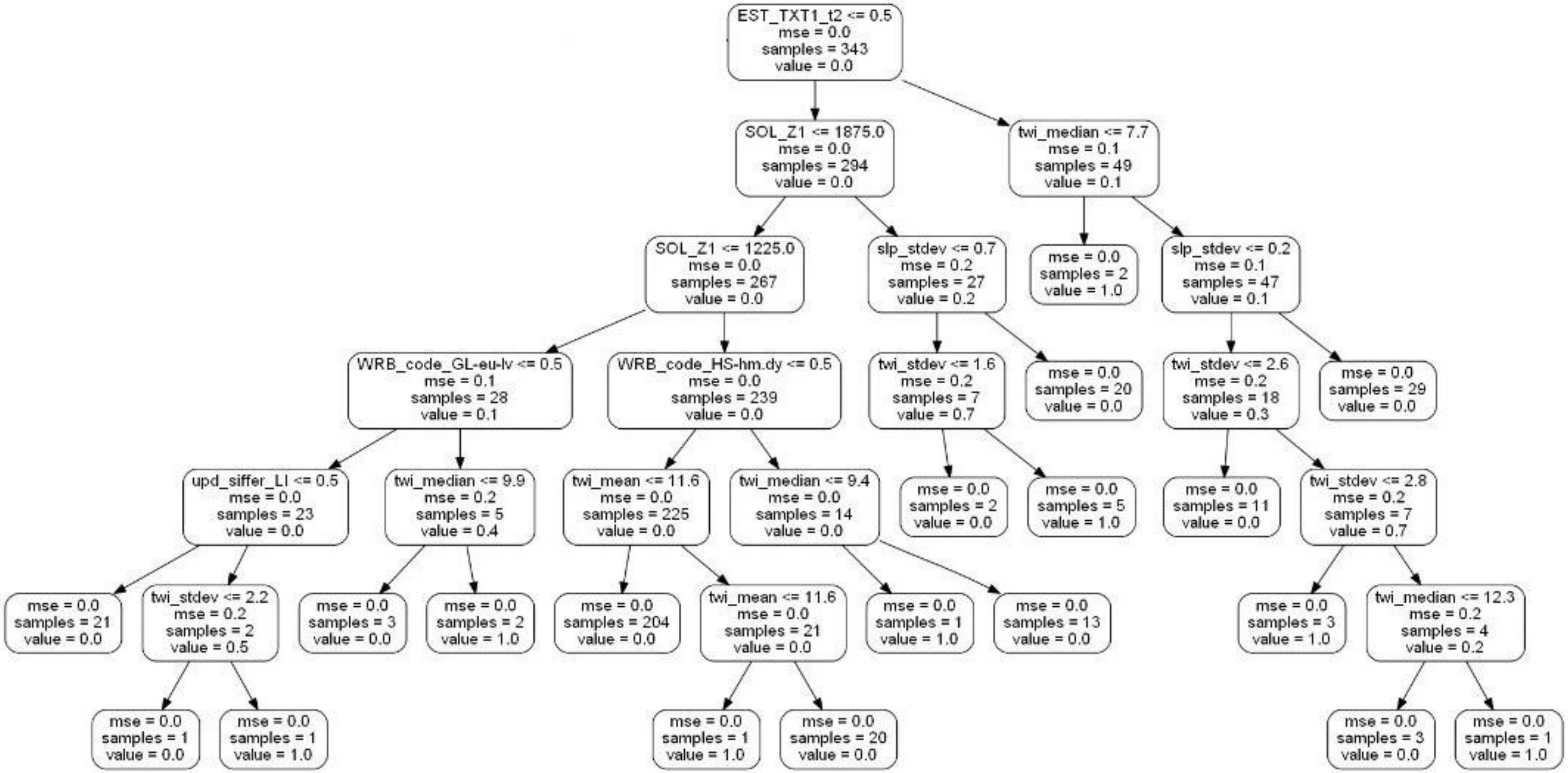


Figure 10. *Example of a single decision tree in the RF model*

Once all the models were run, their prediction statistics were checked. In order to choose the best performing model for further analysis, R^2 values and OOB scores were examined. There were almost no differences between these figures across all the models (Table 10). However, only one model (Model 5) was slightly differentiable from the rest. It contained soil type, soil texture, slope, topographic wetness index and bedrock as predictor variables. Further, split option 2 (70/30) was chosen as suitable splitting option for this particular question. Accuracy of the selected model was 0.79 and 0.8 for R^2 and OOB score respectively. This means that RF reached nearly good prediction results.

During the process, it was possible to check which predictor variable contributes the most to the model or how the accuracy will decrease if a certain variable will be removed from the model. It was calculated via permutation feature importance function. The permutation feature importance is defined to be the decrease in a model score when a single feature value is randomly shuffled (rfpimp) (Breiman, 2001). Eventually, there were nearly similar importance scores for variables in each model, except for bedrock which had significantly higher importance (Figure 11).

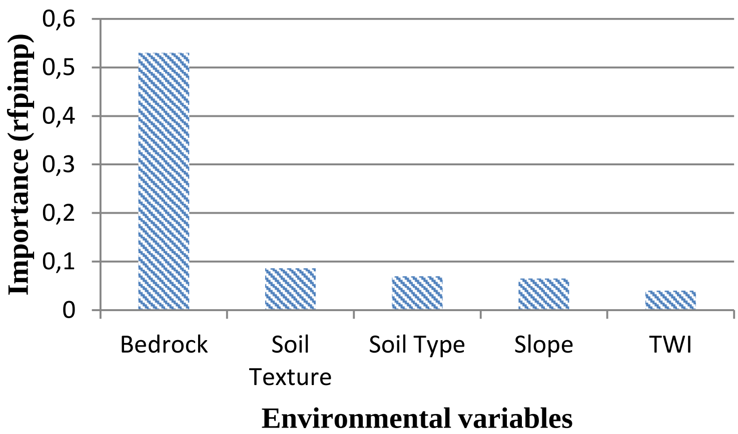


Figure 11. *Importance of each variable in the final RF model*

Table 10. *Statistics of the different RF models tested in the study*

Variant	Test split	R^2 score	OOB score	Pearson's test	Pearson's train
Model 1	1	0.7617	0.75438	0.88126	0.98733
Model 2	1	0.7457	0.74309	0.87174	0.98671
Model 3	1	0.7888	0.75424	0.89932	0.98750
Model 4	1	0.7783	0.77606	0.89119	0.98857
Model 5	1	0.7631	0.76548	0.88155	0.98777
Model 1	2	0.7756	0.77503	0.88934	0.98862
Model 2	2	0.7884	0.79678	0.89560	0.98967
Model 3	2	0.7919	0.79670	0.89547	0.98880
Model 4	2	0.7861	0.79668	0.89403	0.98958
Model 5	2	0.6692	0.68487	0.82575	0.98546

The predicted suitability ranged from 0.04 to 0.884 %. Higher than 80% suitability probability was considered as highly suitable and less than 80% was considered as not or low suitable for alvars. The aim was to find very high probability suitable areas and therefore high threshold was selected.

From all the suitable areas for alvars, 45% were currently forests, 34% are croplands and 11% are grasslands (Table 11).

Table 11. Actual land use under the predicted suitable areas for alvar restoration with RF

Landuse	**Areal percentage (%)**
Forest	44.94
Cropland	33.49
Grassland	11.02
Other	6.66
Shrubland	2.32
Urban	1.14
Wetland	0.34
Water	0.09

As a result, RF predicted 610.91 km^2 of suitable areas where nowadays no alvars exist on (Figure 12). From those, 470 km^2 were once alvar areas in the past. The most suitable areas appeared in the western islands of Estonia (Saaremaa, Muhu, Hiiumaa) as well as north-west inland areas and northern Estonia. Southern Estonia fell into low or no suitability areas.

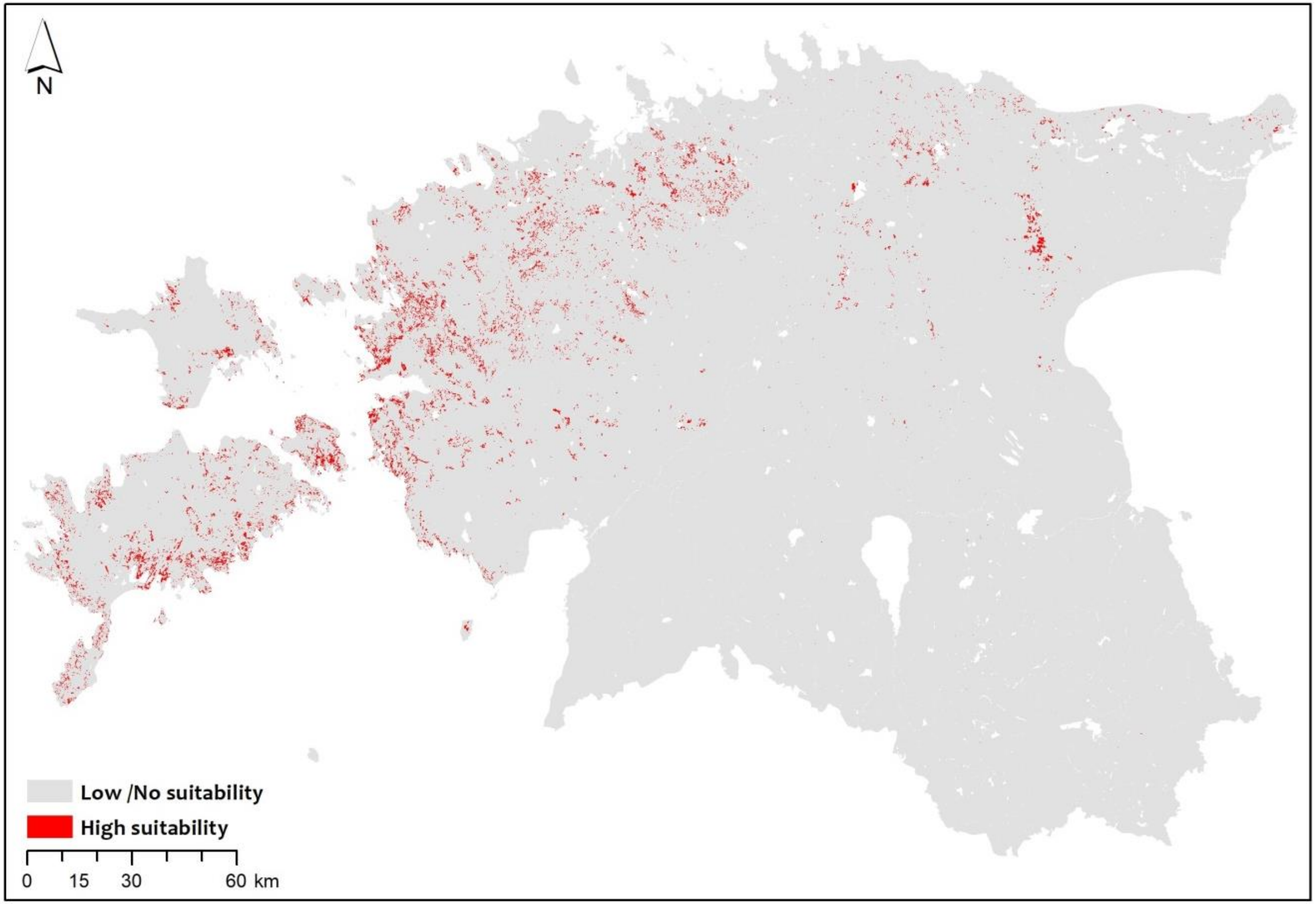


__Figure 12.__ Results of RF model in land suitability analysis of alvars in Estonia

5. Discussions

Based on the literature review, one of the most important variables in alvar identification was considered to be the soil depth since Albert (2006) characterizes alvars with thin soils up to 20 cm. Therefore, variation of soil depth in the soil database was inspected. It became clear that the available soil depth information from the soil database cannot be utilised. Because the soil depth has not been measured consistently throughout Estonia. Some areas had an actual soil depth record while others were showing only the depth of a soil profile up to 1 m. depth.

Pärtel (1999) distinguished various alvars based on the state of the soils they are occurring on. "Dry alvars", according to him, are on shallow soil layer that dries out and results in the drop of vegetation. "Wet alvars" are formed as a result of occasional and/or sometimes permanent water pools happening during rainy seasons. When examining soils under the alvars, this distinction was clearly noticeable. It was possible to clearly distinguish "dry" alvars that are occurring on soil types Rendzik Leptosols and Rendzik Leptosols + Calcaric Regosols as well as "wet" alvars occurring on soil types Rendzi- Gleyic Leptosols and Rendzi-Gleyic Leptosols + Calcari Gleyic Regosols. Most of the alvars occurred on sand clay and clay sand and clay textures.

As Helm et. al., (2007) explained, alvars are occurring on the areas limestone bedrock outcropping occurs which was strongly confirmed by the data used in this study.

Land use change by itself is not a measure or parameter for alvar restoration activities but is rather an important trigger for consideration of restoration. According to Pärtel (1999), change of land management practices because of land use change, resulted in alvars being overgrown by shrubs and trees. Indeed, nowadays alvars have significantly been altered into different land uses if compared with data from 1930ties. Since the overgrowth is one of the signs of alvar change, it is not surprising that from all the land uses to which alvars turned into, forest is on the first place.

MCDM has been used in many studies for mapping habitat suitability and it has shown good results (Uuemaa et. al., 2018). Therefore, it was expected that with this method, highly accurate results will be achieved. However, MCDM is strongly relying on expert knowledge and it can be subjective. Thus, standardisation or definition of weights for the criteria requires strong attention (Romano et. al., 2015). In this study the subjectivity is, however, minimized by using AHP method.

By using the expert knowledge based pairwise comparison matrix, weights for all the environmental variables were calculated. This means that during the weighted overlay analysis, decisions were made based on the weights of these variables and it followed the following order (descending): bedrocks, soil type, land use, slope, TWI and soil texture. This order of decision-making process can actually explain the results. For instance, highly suitable areas were mostly found in the western and the northern Estonia, since alvar suitable bedrocks are located in these areas. Because all the land use classes received weights and none of them was claimed as completely unsuitable then some areas of "not suitable" classes (e.g. water or wetland) got high suitability if they were occurring on very suitable bedrock.

The accuracy of predictions was assessed only using known locations of alvars that nowadays exist. First, four classes based accuracy was calculated. 98.7% of nowadays existing alvars were identified. Afterwards, the accuracy of "highly suitable" class was checked. The class 4 identified only 27.4 % of nowadays existing alvars.

In fact, assessment of accuracy of predictions based on only existing alvars' location might not reflect the suitability the best. It might be that alvars have been forced out from their most suitable locations and still exist in the areas that are not so ideal to them. However, from the conservation perspective, if they can strive in these not so suitable areas and the species richness is still high and there is not so high pressure by humans into these areas, then these might be worth areas to consider for alvar conservation.

Considering the available datasets and their limitations, the need for more independent technique emerged. Thus, RF was chosen to be used in this thesis work. There are already many examples of the application of RF in suitability assessment. Further, RF has ability to learn by itself from the available data (in this case current locations of alvars) and make predictions based on that without human restriction. Considering that it is quite difficult to do analysis and make prediction in country level, RF was expected to give good results.

In order to find best set of variables for suitability assessment, different models were built and the best among them was chosen. As it was expected from the beginning, bedrock, soil texture and soil type had the highest contribution to the final model and helped the most to precisely identify suitable areas. In many similar studies where RF has been used for habitat suitability analysis, performance of the model above 90% was usually considered as acceptably accurate. For instance, in their habitat suitability analysis using machine learning techniques, Benito et. al. (2006) reached 98% accuracy for RF model. Although the accuracy of predictions in this work was around 80%, the results were thoroughly examined and the experts of UT have approved that the results are reasonably good. RF was learning the data and building relationship between the data by itself. Therefore, RF predicted in comparison to MCDM slightly more areas to be suitable for alvar restoration on actually not suitable land uses such as wetland or urban areas. However, this difference is minimal and is not enough to conclude that one method performing better over another one.

Eventually, there was approximately 400 km^2 difference in prediction of suitable areas for alvar restoration between two methods, where MCDM predicted more than RF. This number can be explained by the overall conceptual differences between these methods. For MCDM all the datasets involved into analysis were reclassified beforehand, based on the weights of criteria. Further, MCDM considered all the areas in Estonia when assessing potential suitability. While only the areas where alvars existed/exist were used to train RF models, and these were not very big datasets. Therefore, it was predicting existence of alvars or their current/historic locations. Thus, absence of alvar in certain area does not mean that this area is not suitable for alvar. And therefore, accuracy of RF was higher and less areas have been predicted.

All in all, MCDM predicted suitable areas for alvar restoration or creation of alvar-like habitat both where alvars actually exist nowadays and elsewhere, where there are no alvars present. RF predicted mostly areas for alvar restoration on known alvar locations. Since potentially suitable areas were identified based on the environmental variables, all the predicted areas need to be validated using different means, for instance soil and/or topographic survey.

Previous alvar restoration practices were mainly based on the Swedish experience, where since the beginning of 90's approximately 7000 ha of alvars have been restored. Additionally from recent large-scale grassland restoration activities in the project LIFE to Alvars (LIFE13 NAT/EE/000082) (Holm, 2019) ca 3000 ha alvar grasslands have been restored. However, the restoration has been focusing on the existing alvars, while the potentially suitable environmental conditions might allow to create alvar-like habitats also in the regions that have not historically nor currently been alvar grasslands. This would increase the available area for species dependent on alvar habitat. With the results acquired from this thesis, it might be possible to extend the scale of the restoration activities and cover considerably more areas than the existing alvar areas. In land suitability analysis the priority land use for alvar restoration was chosen grasslands and scrublands while forests were the second most suitable areas for restoration. If the previous restoration works were aiming immediate clean-up of juniper shrub coverage in existing alvars in order to prevent further overgrowth of these grasslands (Helm, 2006), then a more collaborative and detailed framework organisation for alvars restoration/creation might be necessary in the new circumstances. It is also important to consider competing land-use options and further limit alvar-like habitat restoration/creation to areas that do not hold other conservation values.

It is said that the quality of the planning process directly depends on the availability of the data and existence of a proper and reliable data processing tools. The better data processing is the better planning results will be (Hall, 1974). Therefore, it is suggested that for actual restoration planning, the whole process must be repeated with the improved datasets. Especially, the soil database should be enriched with actual soil depth data. Furthermore, additional attributes such as soil pH etc. might be useful in proper site selection for alvar restoration. Further, alvar restoration practices based on the given results must consider the spatial configuration of the selected sites. Habitat fragmentation and decreasing patch sizes drastically influence species richness within the patches or sites (Butaye et al., 2005). Therefore, predicted suitable areas with the biggest area and/or with the highest number of neighbouring suitable patches should be considered for further restoration activities. In this way persistence of the alvar grasslands might be achieved.

Concluding, both methods gave considerably high results in the land suitability analysis. Therefore, it is difficult to compare these two methods based on their results for alvar suitability assessment. However, it is possible to compare them based on the workload and dependency on external factors. Thus, RF method is suggested to use, if there are no experts on the studied topic available and it is difficult to establish which environmental variables will describe the problem in a best way. In this case, RF will learn by itself the relationship between variables and their spatial distribution and will make predictions based on the learned information. Further, RF might tremendously reduce the time contributed to the suitability analysis but will require more detailed examination of the results. MCDM together with AHP is a very well-known and the most frequently used technique to find suitable areas. Because of the possibility to control all the stages of the analysis, MCDM might give better results. However, this process is very time consuming and requires good knowledge and/or availability of the experts on the studied topic.

In the case of this master thesis, RF and MCDM helped to achieve the overall research aim. As such, (1) totally new areas where alvars never existed before but the combination of different environmental parameters proved these areas to be suitable for creation of alvar-like vegetation, and (2) the areas once historically covered by alvars but which were lost due to the heavy human intervention and change of land use practices were identified.

6. Conclusion

In this thesis, the importance of the alvar grasslands in Estonia was underlined. Their species richness, their capability to contribute to the ecosystem and their socio-economic potential is a very strong reason to consider their protection and restoration. The reduction trend of area of alvars and the conversion of those areas to different land uses such as forests, shrubs and etc. has led to the necessity of restore these habitats. Therefore, using two different methods, environmentally potentially suitable areas for alvar restoration were identified. Various datasets, including alvar distribution data, Soil database, Digital Topographic database and Land use data for the whole Estonia were utilized. Before proceeding to land suitability analysis, preliminary statistical and spatial analysis on the existing datasets were performed. This process helped to understand the general picture behind the datasets. This way it was established that there are particular soil types (Rendzik Leptosols, Rendzik Leptosols + Calcaric Regosols, Skeletic Leptosols, Calcari Abruptic Gleysols), textures (clay sand, sand clay) and bedrocks (Cambrian, Ordovician and Silurian) that alvars occur on are included and suitably considered in the analysis. Furthermore, the range of slopes (0-1.5) and topographic wetness indexes (8-11) that were found to characterise current and historical alvars were identified. Lastly, the land uses to which alvar areas are turned into were clarified: forest, agricultural fields and woody areas are the most common “occupiers”.

In order to find environmentally suitable areas for alvar restoration in Estonia, within and outside of their common occurrence area, Multi Criteria Decision Making approach and RF model of Machine learning technique were used.

A very well-known method in land suitability analysis, Multi-criteria decision making approach together with the Analytical Hierarchy process was implemented to the alvar suitability analysis. For this process pairwise comparison matrices for the criteria and sub-criteria were built. The criteria and sub-criteria were formed from the following variables: soil type, soil texture, bedrock, slope, topographic wetness index and land use. Experts of the botany department of University of Tartu evaluated the tables and rated each of the criteria and the sub-criteria based on the importance table by Saaty (2008). Derived from the given values, consistency ratio as well as weights for each criterion were calculated. Further, weights of each criterion formed the basis of reclassification process of the used layers while weights of each criterion helped to prioritize these criteria during the weighted overlay analysis. Once again, the results of the weighted overlay analysis were evaluated by the experts of the botany department. Finally, using the AHP incorporated to the MCDM approach in alvar suitability analysis, 987.93 km^2 of highly suitable areas for alvar restoration have been identified.

RF is an ensemble method of decision trees used for both regression and classification tasks. Different RF models were created using various base set of parameters and variables. Afterwards, the best performing model among all the others was selected and was later used to assess the probability of areas being suitable for alvar restoration. Although soil depth is one of the most important variables that characterize alvars, it was eliminated from the predicting model. Soil depth information was not consistently sampled throughout Estonia and thus was distorting the accuracy of the predictions. The best performing model was run on the final dataset for the whole Estonia. Results were plotted and evaluated by the experts of the botany department of the University of Tartu. Eventually, RF predicted 610.91km^2 of areas having high probability of being environmentally suitable for restoration of alvar grasslands and creation of alvar-like vegetation.

In conclusion, land suitability analysis can help to reveal areas where certain habitats can be or cannot be restored or created by using novel methods.

Both methods resulted in the reasonably accurate predictions. Although from results it is difficult to prioritize one method over another one in land suitability analysis for alvars, few further suggestions might improve the results:

1. Improvement of datasets. Although using available datasets it was possible to reach good results, including additional information such as soil depth, soil pH content might increase the performance of the both methods
2. With the existing dataset, for MCDM approach suitability classes need to be clearly defined. The threshold that defines when an area is considered as low/highly suitable determines how accurate/noisy results will be in the end.
3. Spatial configuration and the number of the neighbouring suitable areas must be discussed. The best areas, in terms of fewer fragmentation and higher number of neighbour areas, should further be considered for restoration.

Summary

Suitability Analysis for Alvars in Estonia using Random Forest and GIS based Multi Criteria Decision Making approach

Irada Ismayilova

Alvar grasslands are biodiverse habitats where dispersed shrubs and rare tree coverage occurs. Alvars are flat, relatively open areas with shallow or sporadic soil cover (often < 20 cm) over calcareous limestone or dolomite bedrock (Albert, 2006). There is limited distribution of alvars in the world. This makes alvars globally rare and emphasizes the need of their protection (Helm, Urbas, & Pärtel, 2007). In Estonia, they are mostly found in Saaremaa, Muhu, Läänemaa, Hiiumaa, as well as in Harjumaa, Ida and Lääne-Virumaa.

In Estonia, alvars originated and developed under human influence within thousands of years, where human management, such as cutting hay or animal grazing (sheep, horses) and deforestation was the reason of grasslands' persistence (Pärtel et al., 1999). Due to high amount of vascular plat species, in Estonia alvars are considered to be second species rich communities after wooded meadows (Helm, 2006). They were primarily used as pastures. Currently, because of the significant change in pattern of traditional land use practices, in Estonia alvars are no longer used for grazing and they are overgrown with shrubs, trees and tall grazing-sensitive herbaceous plants.

Up to now, restoration of alvars is mainly done on existing alvars by reducing areal coverage of unwanted vegetation and by introducing proper land use management techniques ("LIFE to alvars"). This thesis will focus on other methods which will give an input for future restoration practices of alvars.

Land use suitability analysis is one of the most frequently used techniques in environmental management. Simple premise of land use suitability analysis is that in any case there are environmental characteristics which are either suitable or unsuitable for the planned activity of analysed situation (Parry et al., 2018). With increasing urge of habitat conservation, growing number of literature also became available on land suitability analysis of different land uses for protection and conservation purposes.

Therefore, RF method of machine learning technique and GIS based land suitability analysis, together with Multi Criteria Decision Making approach will be used for land suitability analysis. Areas suitable for restoration of alvars in whole Estonia will be searched based on the combination of different environmental variables. The environmental datasets, used to derive environmental variables, consisted of a soil database, a digital elevation model (DEM) and Estonian Digital Topographic Database.

RF is a decision tree based classifier and can be described as trees, where branches formed by the answers to yes/no questions and are not pruned. Each tree in the forest constructed using bootstrap sample from the original dataset. It uses random selection of explanatory variables or factors to split the tree at nodes, instead of splitting each node based on the best split among all the variables, thus avoiding overfitting. The goal of RF is to identify the best model to analyse the relationship between dependant and independent variables (Friedman et.al., 2003).

MCDM is a process of finding the best alternative from the set of relevant alternatives (Sánchez-Lozano, 2013). The process of the MCDM approach consists of several interdependent steps which result in the final decision on the studied topic (Pohekar & Ramachandran, 2004). The most frequently chosen method from the various MCDM methods is analytical hierarchy process (AHP). AHP manages criteria into a hierarchy tree

where the upper level is the problem to which a solution is looked for and the lowest level contains various sub criteria or parameters.

One of the most important variables in alvar identification was considered to be the soil depth. However, after studying the soil database of Estonia, unfortunately, it was established that soil depth information cannot be utilised, due to its inconsistent sampling throughout Estonia. Furthermore, preliminary statistics on the alvar distribution datasets showed that historically alvars mostly occurred on Rendzik Leptosols (Kh), Rendzik Leptosols + Calcaric Regosols(K), Skeletic Leptosols (Kr), Calcari Abruptic Gleysols (Gh) and Rendzi- Gleyic Leptosols (Khg). Whereas existing nowadays alvars are distributed mostly on the Leptosols+calcaric Regosols (K), Skeletic Leptosols (Kr), Rendzi-Lithic Leptosols (Kh), Rendzi-Gleyic Leptosols + Calcari Gleyic Regosols(Kg), Mollic Gleysols (Go). Similar to soil types there was a clear preference to one soil texture by historical alvars over another soil texture by nowadays existing alvars and sand (SL), sandy and clay sand (SL) were the most common textures. From the available alvar distribution datasets both in the past and nowadays, alvars were located on the Silurian, Cambrian and Ordovician bedrocks. Statistics showed that slopes of 0-1.5 degrees are the most optimal for alvar occurrence while topographic wetness index of 8-11 were the most suitable and indices between 0-8 and 11- 23 were mostly unsuitable. Statistics revealed that there was a significant change in the land use under alvars. Hence, 48% of alvar territories were lost to forests, 11% of alvar areas are occupied by agricultural fields and 10% by shrublands. Approximately, 3% is occupied by wetlands, 2% of initial alvar areas are now occupied by private lands, 1% by water bodies and 10% by other land uses. And only 15% of the area remained grasslands.

RF model predicted 610.91 km^2 of highly suitable areas for alvar restoration. MCDM method predicted 987.93 km^2 areas for alvar restoration. Results of both methods have been validated. MCDM reached 27.4% of accuracy in the case of having two suitability classes and 98.7% in the case of having four suitability classes. RF prediction accuracy was 80%. Although, there was not too big difference between predictions made by both methods, under the existing circumstances, the more suitable areas for restoration of alvars or creation of alvar-like habitat exist, the more successful the restoration/creation process at the end will be.

Finally, following suggestions were made:

1. Datasets could be improved. Although using available datasets it was possible to reach good results, including additional information such as soil depth, soil pH content might increase the performance of the both methods
2. With the existing dataset, for MCDM approach suitability classes need to be clearly defined. The threshold that defines when an area is considered as low/highly suitable determines how accurate/noisy results will be in the end.
3. Spatial configuration and the number of the neighbouring suitable patches must be discussed. The best areas, in terms of fewer fragmentation and higher number of neighbour areas, should further be considered for restoration.

Kokkuvõte

Eestis alvarite taastamiseks sobilike maade analüüs, kasutades otsustumetsa ja GIS-il põhineva mitmekriteeriumilise otsustusanalüüsi meetodit

Irada Ismailova

Alvarid on bioloogiliselt mitmekesised elupaigad, kus on levinud põõsas- ja puhmastaimed ning hajusalt kasvavad puud. Alvarid on tasased, suhteliselt avatud alad õhukese mullakattega (sageli < 20 cm) lubjarikkal paekivist või dolomiidist aluskivimil (Albert, 2006). Alvarite levik maailmas on piiratud. See muudab alvarid kogu maailmas haruldasteks ning rõhutab veelgi nende kaitsmise vajadust (Helm, Urbas ja Pärtel, 2007). Eestis leidub neid peamiselt Saaremaal, Muhu saarel, Läänemaal, Hiiumaal, aga ka Harjumaal ning Ida- ja Lääne-Virumaal.

Eestis tekkisid ja arenesid loopealsed inimmõjul tuhandete aastate kestel paigus, kus rohumaade püsivuse põhjuseks oli inimtegevus, näiteks heina niitmine või loomade (lammaste, hobuste) karjatamine ja raadamine (Pärtel et al., 1999). Soontaimeliikide suure arvukuse tõttu loetakse alvareid Eestis liigirikkuse poolest teiseks koosluseks puisniitude järel (Helm, 2006). Neid kasutati peamiselt karjamaadena. Praegu Eestis loopealseid oluliste muutuste tõttu maakasutuses karjatamiseks enam ei kasutata ning nad on põõsaid, puid ja karjatamise suhtes tundlikke kõrgekasvulisi rohttaimi täis kasvanud.

Seni on loopealseid taastatud peamiselt olemasolevatel loopealsetel, vähendades soovimatu taimestiku katvust ning rakendades sobivaid maakasutusvõtteid (programmi „ELU loopealsetele“ raames). Selles töös keskendutakse muudele meetoditele, mis annavad sisendi loopealsete edaspidisteks taastamisvõteteks.

Maakasutuse sobivusanalüüs on keskkonnahalduses üks sagedamini kasutatavaid tehnikaid. Maakasutuse sobivusanalüüsi lihtne eeldus on, et igal juhul on olemas keskkonna omadused, mis analüüsitavas olukorras kavandatud tegevuseks sobivad või mitte (Parry et al., 2018). Elupaikade säilitamisvajaduse suurenemisega seoses on üha rohkem ka maakasutuse sobivuse analüüsi puudutavat kirjandust erinevate maakasutusviiside kohta kaitse ja säilitamise eesmärkidel.

Seetõttu kasutati maa sobivuse analüüsimisel masinõppe meetodit osustusmets ja GIS-põhist maa sobivusanalüüsi koos mitmekriteeriumilise otsustusanalüüsi meetodiga. Loopealsete taastamiseks sobivaid alasid otsiti üle Eesti erinevate keskkonnamuutujate kombineerimise põhjal. Keskkonnamuutujate tuletamiseks kasutatud keskkonnaandmete kogum koosnes mullastiku andmebaasist, digitaalsest kõrgusmudelist (DEM) ja Eesti digitaalsest topograafilisest andmekogust (ETAK).

Otsustusmets kujutab endast otsustuspuu põhist klassifikaatorit, mida võib kirjeldada puudena, mille oksad moodustuvad vastustest jah/ei küsimustele.. Iga puu konstrueeritud metsas kasutab algandmestiku alglaadimisnäidist. Puu tükeldamiseks sõlmkohtadel kasutab see seletavate muutujate või tegurite juhuslikku valikut, selmet jagada iga sõlme parima jaotusega kõigi muutujate vahel, vältides sellega ülesobitamist (overfitting). Otsustusmetsa eesmärk on selgitada välja parim mudel sõltuvate ja sõltumatute muutujate vaheliste seoste analüüsimiseks (Friedman et.al., 2003).

Mitmekriteeriumilise otsustusanalüüs (ingl. k. multiple criteria decision analysis, MCDM) on protsess, mille käigus leitakse asjakohaste alternatiivide hulgast parim (Sánchez-Lozano, 2013). MCDM lähenemisviisi protsess koosneb mitmest üksteisest sõltuvast etapist, mille järel tehakse lõplik otsus. Erinevatest MCDM-meetoditest kasutatakse kõige sagedamini analüütilist hierarhilist otsustusprotsessi (AHP). AHP haldab kriteeriume hierarhiapuuna, kus kõrgem tasand on probleem, millele lahendust otsitakse, ning madalam tasand sisaldab erinevaid alakriteeriume või parameetreid.

Üheks olulisimaks muutujaks loopealsete tuvastamisel peeti mulla tüsedust. Pärast Eesti mullaandmebaasi uurimist kahjuks siiski tõdeti, et mulla tüsedust puudutavat teavet kasutada ei saa, kuna andmed ei ole ühtlase kvaliteediga. Loopealsete jaotuse andmekogumite esialgne statistika näitas, et ajalooliselt on loopealsed esinenud peamiselt paepealsetel muldadel (Kh), paepealsetel muldadel + lubjarikastel erodeeritud muldadel (K), koreserikastel rähkmuldadel (Kr), õhukestel paepealsetel gleimuldadel (Gh) ja gleistunud õhukestel paepealsetel muldadel (Khg). Tänapäeval paiknevad olemasolevad loopealsed enamasti paepealsetel muldadel + lubjarikastel erodeeritud muldadel (K), koreserikastel rähkmuldadel (Kr), õhukestel paepealsetel muldadel (Kh), gleistunud karbonaatsetel muldadel + lubjarikastel gleistunud rähkmuldadel (Kg) ja leostunud gleimuldadel (Go). Sarnaselt mullatüüpidega võis ajalooliste loopealsete korral märgata selget eelistust kindlale lõimisele võrreldes tänapäevaste loopealsetega; kõige levinumad lõimised olid liiv-, saviliiv- ja liivsavi. Nii varasemate kui tänapäevaste saadaolevate loopealsete levikuandmestike põhjal paiknesid loopealsed siluri, kambriumi ja ordoviitsiumi aluskivimitel. Statistika näitas, et loopealsete esinemiseks on kõige optimaalsem maapinna kalle 0–1,5 kraadi ja sobivaim topograafiline niiskusindeks 8–11; indeksid 0–8 ja 11–23 olid enamasti sobimatud. Tulemused näitasid, et loopealsete maakasutus on oluliselt muutunud. Nii on 48% loopealsetest aladest kattunud metsaga, 11% on põldude ning 10% põõsastike all. Ligikaudu 3% loopealsetest on nüüdseks märgalad, 2% kunagistest loopealsetest on eramaad, 1% veekogude all ja 10% on hõlmatud muu maakasutusviisiga. Ainult 15% loopealsete aladest on säilinud rohumaadena.

Otsustusmetsa prognoosi kohaselt on loopealsete taastamiseks väga sobivaid alasid 610,91 km^2. MCDM-meetodi alusel prognoositi, et neid alasid on 987,93 km^2. Mõlema meetodi tulemused on valideeritud. MCDM saavutas kahe sobivusklassi kasutamisel 27,4%-se täpsuse ning nelja sobivusklassi kasutamisel 98,7%-se täpsuse. Otsustusmetsa prognoosi täpsus oli 80%. Ehkki kummagi meetodi alusel tehtud prognooside vahel ei olnud liiga suuri erinevusi, siis olemasolevates tingimustes on loopealsete või loopealselaadsete elupaikade taastamisel parem, kui leitakse selleks võimalikult sobivad alad – seda edukam saab olema taastamise või loomise lõpptulemus.

Lõpuks tehti järgmised ettepanekud.

1. Andmebaase võiks parandada. Ehkki ka olemasolevaid andmebaase kasutades oli võimalik jõuda heade tulemusteni, võiks lisateabe, näiteks mulla tüsedust ja pH-sisaldust puudutavate andmete olemasolu mõlema meetodi tulemuslikkust suurendada.
2. Olemasoleva andmebaasi korral tuleb MCDM-metoodikat kasutades selgelt määratleda sobivusklassid. Künnis, millega määratletakse ala vähe või väga sobivana, määrab selle, kui täpsed või segased tulemused lõpuks saadakse.
3. Läbi tuleks arutada ruumiline konfiguratsioon ja naabruses asetsevate sobivate maatükkide arv. Väiksema killustatuse ja suurema naaberalade arvu poolest parimaid alasid tuleks taastama asumisel täiendavalt hinnata.

Acknowledgment

I would like to express sincere gratitude to my supervisor, Dr. Evelyn Uuemaa, who gave me the idea of the thesis topic and helped me throughout the work. I am thankful to Dr. Aveliina Helm, for consulting me in all the important questions related to alvars in Estonia. Special thanks to Dr. Alexander Kmoch for his help with programming tasks and for the contribution to the success of my thesis. I am grateful to Dr. Age Poom for her understanding and all the help throughout my Master's studies. Finally, yet importantly, I want to thank all my friends and my family in Azerbaijan, Estonia and Germany for their understanding and support during my studies.

List of references

Ahmed, G. B., Shariff, A. R. M., Balasundram, S., K. & Fikri Bin Abdullah, A., 2016. Agriculture land suitability analysis evaluation based multi criteria and GIS approach. IOP Conference Series: Earth and Environmental Science 37

Albert, A. 2006. Natural community abstract for alvar. Michigan Natural Features Inventory, Lansing, MI. 10

Benito, G., Blazek, R., Neteler, M., Dios, R., Ollero, H. and Furlanello, C. 2006. Predicting Habitat Suitability with Machine Learning Models: The Potential Area of Pinus sylvestris L. in the Iberian Peninsula. Ecological Modelling 197:383-393

Borcherding, K., Schmeer, S., Weber, M., 1995. Biases in multiattribute weight elicitation. (ed.) J-P. Caverni. Contributions to Decision Making. Amsterdam: Elsevier.

Breiman, Leo, 2001. Random forests. Mach. Learn. 45 (1), 5–32. Chen, X., Ishwaran, Hemant, 2012. Random forests for genomic data analysis. Genomics 99 (6), 323–329

Buchanan, B. & Fleming, M.L. & Schneider, Rebecca & Richards, Brian & Archibald, Josephine & Qiu, Zeyuan & Walter, M.. (2013). Evaluating topographic wetness indices across Central New York agricultural landscapes. Hydrology and Earth System Sciences Discussions. 10. 14041-14093. 10.5194/hessd-10-14041-2013

Butaye, J., Adriaens, D., Honnay, O., 2005. Conservation and restoration of calcareous grasslands: a concise review of the effects of fragmentation and management on plant species. Biotechnology, Agronomy, Society and Environment 9: 111-118

Chandio, I. A., Matori, A. N. B., WanYusof, K. B., Talpur, M. A. H., Balogun, A. L. & Lawal, D. U., 2013. Arabian Journal of Geosciences 6: 3059-3066

Chang, N.B. & Pires A., 2015. Sustainable Solid Waste Management: a Systems Engineering Approach. Wiley, New York, USA

Chen Y, Yu J, Khan S, Environmental modelling and software, 2010, 25, 1582– 1591

Cheng, S., Chan, CW., Huang, GH., 2003. An integrated multi-criteria decision analysis and inexact mixed integer linear programming approach for solid waste management. Engineering Applications of Artificial Intelligence 16(5):543–554

Church, R. L., 2002. Geographical information systems and location science. Computers and Operations Research 29:541–556

City, L., 2011. GIS-based Land Suitability Analysis Using AHP for Public Parks Planning in GIS- based Land Suitability Analysis Using AHP for Public Parks Planning in Larkana City. Modern Applied Science, Vol. 5, No.4,

Claudia, A. & Douglas, W., 1997. Vegetation, environmental characteristics and ideas on the maintenance of alvars on the Bruce Peninsula, Canada. Journal of Vegetation Science 8: 797-810

Climaco J., 1997. Multicriteria analysis. New York: Springer-Verlag

FAO, 1976. A framework for land evaluation, Soil Bulletin 32. Food and Agriculture Organization of the United Nations, Rome, Italy

FAO, 1976. A framework for land evaluation, Soil Bulletin 32. Food and Agriculture Organization of the United Nations, Rome, Italy

Fernandez, I., & Morales San Martin, N., 2016. A spatial multicriteria decision analysis for selecting priority sites for plant species restnarkioration: A case study from the Chilean biodiversity hotspot. Restoration Ecology 24: 599-608

Fonseca, M.S., Kenworthy,W. J., & Thayer, G.W., 1998. Guidelines for conservation and restoration of seagrass in the United States and adjacent waters. NOAA/NMFS Coastal Ocean Program and Decision Analysis Series, No. 12. NOAA Coastal Ocean Office, Silver Spring, Maryland.

Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. Stat. Med. 22 (9), 1365–1381

Gazol, A. , Tamme, R., Takkis, K., Kasari, L., Saar, L., Helm, A. & Pärtel, M., 2012. Landscape and small scale determinants of grassland species diversity: direct and indirect influences. Ecography 35: 944-951

Geneletti, D., Duren, I., 2008. Protected area zoning for conservation and use : A combination of spatial multicriteria and multiobjective evaluation. Landscape and Urban Planning 85: 97-110

Georgian, S., Anderson, O., Rowden, A. 2019. Ensemble habitat suitability modelling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. Fisheries Research 211 (2019) 256–274

Hall, P., 1974. Urban and Regional Planning, Penguin, Harmondsworth.

Helm, A., Hanski, I., & Pärtel, M.,2006. Slow response of plant species richness to habitat loss and fragmentation. Ecology Letters 9: 72–77

Helm, A., Urbas, P., & Pärtel, M., 2007. Plant diversity and species characteristics of alvar grasslands in Estonia and Sweden. Acta Pythogeogr. Suec. 88: 33-42

Herngren, L., Goonetilleke, A., Ayoko, G.A., 2006. Analysis of heavy metals in road deposited sediments. Analytica Chimica Acta 571: 270–278

Holzkämper, A., Lausch, A. and Seppelt, R. 2006. Optimizing Landscape Configuration to Enhance Habitat Suitability for Species with Contrasting Habitat Requirements. Ecological Modelling 198: 277-292

Hunter, R., Day, J., Shaffer, G., Lane, R.,Englande, A., Reimers, R., Kandalepas, D., Wood, William B., Day, J., Hillmann, E., Bank, E., 2016. Restoration and Management of a Degraded Baldcypress Swamp and Freshwater Marsh in Coastal Louisiana. Water 8: 79-101

Huxel, G., Hastings, A., 1999. Habitat Loss, Fragmentation and Restoration. Restoration Ecology 7: 309-315

Jack, S. W., Drive, J., Suite, A., 2005. Habitat Suitability Analysis : Compensation for Injured Reef in Support of Restoration Planning for the Berman Oil Spill , San Juan , Puerto Rico. Marine Resources Inc

Joerin, F., Thériault, M. & Musy, A., 2001. Using GIS and outranking multicriteria analysis for land-use suitability assessment. International Journal of Geographical Information Science 15: 153–74

Kaar E., 1986. Loometsad ja loodude metsastamine. (Alvarwälder und Aufforschung der alvaren.) Eesti Looduseuurijate Seltsi Aastaraamat 70: 31-38. (In Estonian, with German summary.)

Kalamees, R., Püssa, K., Zobel, K., & Zobel, M. 2012. Restoration potential of the persistent soil seed bank in successional calcareous (alvar) grasslands in Estonia. Applied Vegetation Science 15(2): 208-218

Kmoch, A., Kanal, A., Astover, A., Kull, A., Virro, H., Helm, A., Pärtel, M., Ostonen, I. and Uuemaa, E., 2019: ESSDD - EstSoil-EH v1.0: An eco-hydrological modelling parameters dataset derived from the Soil Map of Estonia, Earth Syst. Sci. Data Discuss., 1–29, doi:10.5194/essd-2019-192, 2019.

Kuussaari, M., Bommarco, R., Heikkinen, R., Helm, A., Krauss, J., Lindborg, R., Ockinger, E., Partel, M., Pino, J., Rodà, F., Stefanescu, C., Teder, T., Zobel, M., Steffan-Dewenter, I. 2009. Extinction debt: a challenge for biodiversity conservation. Trends in ecology & evolution 24 (10): 564-571

Laasimer, L. 1965. Eesti NSV taimkate. Valgus, Tallinn.

Lahssini, S., Lahlaoi, H., Mharzi, H., Bagaram, M., Ponette, Q. 2015. Predicting Cork Oak Suitability in Maˆamora Forest Using Random Forest Algorithm. Journal of Geographic Information System 7: 202-210

Lee C., 2008. The analytic hierarchy process (AHP) approach for assessment of urban renewal proposals. Soc Indic Res 89: 55–168

Malczewski, J., 1999. GIS and Multi criteria Decision Analysis, Wiley, New York

Malczewski, J., 2004. GIS-based land-use suitability analysis: a critical overview. Progress in Planning, 62: 3–65

McHarg, I.L., 1969. Design With Nature, Wiley, New York

Mohit, M.A., Ali M. M., 2006. Integrating GIS and AHP for land suitability analysis for urban development in a secondary city of Bangladesh. Jurnal alam Bina 8: 1–19

Moreno, D., Seigel, M., 1988. A GIS approach for corridor siting and environmental impact analysis. GIS/LIS'88. Proceedings from the third annual international conference, San Antonio, Texas 2: 507–514.

Murray, T., Rogers, P., Sinton, D., Steinitz, C., Toth, R., Way, D., 1971. Honey Hill: a systems analysis for planning the multiple use of controlled water areas. 2 vols. Report nos. AD 736 343 and AD 736 344. National Technical Information Service, Springfield, Virginia

Novak, B., Short, T., 2000. Creating the Basis for Successful Restoration: An Eelgrass Habitat. Ecological Engineering 15: 239-252

Ouyang, N. L., Lu, S. L., Wu, B. F., Zhu, J. J., Wang, H., 2011. Wetland restoration suitability evaluation at the watershed scale. A case study in upstream of the Yongding River. Procedia Environmental Sciences 10: 1926-1932

Park, S., Céréghino, R., Compin, A. and Lek, S. 2003. Applications of Artificial Neural Networks for Patterning and Predicting Aquatic Insect Species Richness in Running Waters. Ecological Modelling 160: 265-280

Parry, J., Ganaie, S., Sultan Bhat, M., 2018. GIS based land suitability analysis using AHP model for urban services planning in Srinagar and Jammu urban centres of J&K, India. Journal of Urban Management 7: 46-56

Pärtel M., Kalamees R., Zobel M. & Rosén E, 1999. Alvar grasslands in Estonia: variation in species composition and community structure. Journal of Vegetation Science 10: 561-570

Pärtel M.,Mändla R. & Zobel M., 1999. Landscape history of a calcareous (alvar) grassland in Hanila, western Estonia, during the last three hundred years. Landscape Ecology 14: 187-196

Peet, R. K., van der Maarel, E., Rosén, E.,Willems, J. H., Norquist, C. and Walker, J. 1990. Mechanisms of coexistence in species- rich grasslands. Bull Ecol Soc Am 71: 283.

Pohekar, S. D. & Ramachandran, M., 2004. Application of multi-criteria decision making tosustainable energy planning—A review. Renewable and Sustainable Energy Reviews 8: 365-381

Puntsag, G., Kristjánsdóttir, S., Ingólfsdóttir, B., 2014. Land Suitability Analysis for Urban and Agricultural Land Using Gis: Case Study in Hvita To Hvita, Iceland. UNU Land Restoration Training Programme

Rahman, M., Sultana, K. R. & Hoque, A., 2008. Suitable sites for urban solid waste disposal using GIS approach in Khulna city. Bangladesh. Proc. Pakistan Acad, Sci, 45

Romano, G., Dal Sasso, P., Trisorio Liuzzi, G., Gentile, F., 2015. Multi-criteria decision analysis for land suitability mapping in a rural area of Southern Italy. Land Use Policy 48: 131–143

Rosén, E., 1982. Vegetation development and sheep grazing in limestone grasslands of south Öland, Sweden. Acta Phytogeogr. Suec. 72: 1-104

Saaty, T., 1986. Axiomatic foundation of the analytic hierarchy process. Management science 32: 841-855

Sánchez-Lozano, J., Teruel-Solano, J., Soto-Elvira, P., Socorro G. 2013. Geographical Information Systems (GIS) and Multi-Criteria Decision Making (MCDM) methods for the evaluation of solar farms locations: Case study in south-eastern Spain. Renewable and Sustainable Energy Reviews 24: 544–556

Spengler, J., Geldermann, S. H., 1998. Development of a multiple criteria based decision support system for environmental assessment of recycling measures in the iron and steel making industry.J. Clean. Prod. 6: 37-52

Steinitz, C., Parker, P., Jordan, L., 1976. Hand drawn overlays: their history and prospective uses. Landscape Architecture 9: 444–455.

Strecht, P., Cruz, L., Soares, C., Moreira, J., Abreu, R. (2015). A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. Proceedings of the 8th International Conference on Educational Data Mining

Tomlin, C.D., 1990. Geographical Information Systems and Cartographic Modelling, Prentice-Hall, Englewood Cliffs, NJ

Uuemaa, E., Hughes, A., & Tanner, C., 2018. Identifying Feasible Locations for Wetland Creation or Restoration in Catchments by Suitability Modelling Using Light Detection and Ranging (LiDAR) Digital Elevation Model (DEM). Water 10(4): 464

Vaidya, O.S. & Kumar, S., 2006. Analytic hierarchy process: an overview of applications. Eur J Oper Res 160: 1–29

Vincenzi, S., Zucchetta, M., Franzoi, P., Pellizzato, M., Pranovi, F., De Leo, G.A. and Torricelli, P. 2011. Application of a Random Forest algorithm to Predict Spatial Distribution of the Potential Yield of Ruditapes philippinarum in the Venice Lagoon, Italy. Ecological Modelling 222: 1471-1478

Wen, L., Ling, J., Saintilan, N. and Rogers, K. 2009. An Investigation of the Hydrological Requirements of River Red Gum (Eucalyptus camaldulensis) Forest, Using Classification and Regression Tree Modelling. Ecohydrology 2: 143-155

Westhoff, V., 1971. The dynamic structure of plant communities in relation to the objectives of conservation. In The Scientific Management of Animal and Plant Communities. Blackwell Scientific Publication, Oxford: 3-14.

Yaakup, A., Zalina, S. & Sulaiman, S., 2004. Integrated land use assessment (ILA) for planning and monitoring urban development. World Conference on Environmental Management. Environmental Management: Facing the Changing Conditions. Bangi,Malaysia

Zarghami, M., & Szidarovszky, F., 2011. Multicriteria analysis applications to water and environment management. Berlin/Heidelberg: Springer

Znamensky, S., Helm, A., & Pärtel, M., 2006. Threatened alvar grasslands in NW Russia and their relationship to alvars in Estonia. Biodiversity and Conservation 15: 1797-1809

Web Sources

Cournapeau, D., 2007. Retrieved from *https://scikit-learn.org/stable/*

EELIS (Estonian Nature Information System). Retrieved from *http://www.eelis.ee/*

Eriksson, M. & Rosén, E. 2008. Management of NATURA 2000 habitats* Nordic Alvar and Precambrian calcareous flatrocks. Retrieved from *https://ec.europa.eu/environment/nature/natura2000/management/habitats/pdf/6280_Nordic_alvar_flatrocks.pdf*

Holm, A., 2019. Report of action C1. Retrieved from *https://life.envir.ee/sites/default/files/pictures/Annex_2_C.1._Restoration_report.pdf*

Estonian Land Board. Retrieved from *https://geoportaal.maaamet.ee/est/Ruumiandmed-p1.html*

LIFE to alvars. Retrieved from *https://life.envir.ee/elualvaritel*

Non-exclusive licence to reproduce thesis and make thesis public

I, Irada Ismayilova,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

"Suitability Analysis for Alvars in Estonia using Random Forest and GIS based Multi Criteria Decision Making approach" supervised by Dr. Evelyn Uuemaa and Dr. Aveliina Helm

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in p. 1 and 2. 4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Irada Ismayilova

Tartu, 16.01.2020