Norwegian University of Life Sciences
Faculty of Environmental Science and Technology
Department of Ecology and Natural Resources Management

Master Thesis 2014
30 credits

Distribution Modeling of Vegetation Types in Venabygdsfjellet, Oppland

Ilya Viktorovich Palkhanov

Preface

This study is the final work of my Master degree in the Natural Resource Management at the Norwegian University of Life Sciences.

I would like to thank my supervisors Senior researcher Anders Bryn and Associated Professor Kari Klanderud for the guidance and valuable remarks and corrections of my master thesis.

I am grateful to Anders Bryn for suggesting the topic of thesis, for assisting me with the fieldwork in Venabygdsfjellet in August 2013, helping with MaxEnt software, and great interest both in the progress and results during studying process. I thank Kari Klanderud for the help with my master thesis, as well as to expand my knowledge of vegetation mapping during the BOT270 course and for well performed mapping fieldwork in the summer 2013.

Finally, I thank Michael Angeloff and Ingvild Nystuen for technical help in ArcMap and obtaining map data. I am grateful to The Norwegian Forest and Landscape Institute for the provided room and technical means. I appreciate "GIS modeling team" at Natural History Museum at University in Oslo for the expansion of my knowledge in the distribution modeling during BIO4115 course.

Norwegian University of Life Sciences
Ås 15th Mai

Ilya Viktorovich Palkhanov

Summary

This study explores the effect of increasing sample units density with presence-only data (PO data) on the ability to predict the distribution of three common (2e – dwarf shrub heath, 4b – bilberry birch forest and 9c – fen) and three rare (3b – tall forb meadow, 8d – rich swamp forest and 9d – mud- bottom fens and bogs) vegetation types.

The chosen study area was Venabygdsfjellet in Ringebu municipality, Oppland. In 2001 the vegetation in the study area was mapped by Norwegian Institute for Forest and Landscape. The vegetation map was used as material for the PO data in the prediction modeling. In beforehand, this map was quality assessed. To evaluate the quality of the map, necessary fieldwork and statistical analysis was conducted. As a result of this evaluation, 84 % of all observations correspond to the mapped distribution on the vegetation map. The PO data for distribution modeling were collected in a point grid with different densities (100 m for common and 25 m for rare vegetation types) within the sample units (1500×600m size). The sample unit was equivalent to a Primary Statistical Unit (PSU) of the AR18×18 survey system and given in a grid net with five densities: 3×3 km, 4,5×4,5 km, 6×6 km, 7,5×7,5 km and 9×9 km. In addition to PO data, 12 environmental variables were used as explanatory predictors (the digital elevation model, basin, curvatures, flow accumulation, flow direction, groundwater, slope, satellite image, the Normalized Difference Vegetation Index (NDVI), the Topographic Wetness index (TWI), sediment and soil maps). Using the PO data and these environmental variables, each vegetation type was modeled in all five densities of the PSU grid using a maximum entropy modeling method using a custom-made software called MaxEnt.

In total, 26 out of 30 planned prediction models were run. The four missing models did not have any PO-points in some of the PSU grid density. Out of 26, 23 prediction models performed well according to the AUC-measure provided by MaxEnt (> 0.80 AUC). The statistical comparison of the predicted and true distribution of the modeled vegetation types showed that only 7 prediction models can be considered as good (2e in densities 3×3 km and 4.5×4.5 km, 4b in densities 3×3 km and 4.5×4.5 km, 9c in densities 3×3 km and 7.5×7.5 km and 3b in density 3×3 km). The vegetation types 8d and 9d were not modeled successfully any PSU grid densities, although they had high AUC-values. The best modeled vegetation type was 4b in a 3x3 km PSU grid density. The variable importance analysis conducted by MaxEnt trough the Jack-Knife test, showed that the DEM (the digital elevation model), NDVI index (the Normalized Difference Vegetation Index), slope and satellite images in blue band were the most important environmental variables among all vegetation type models.

Content

1 Introduction

Vegetation mapping is a very important part of resource management and environmental research both in Norway (Fremstad 1997; Rekdal & Larson 2005) and other countries (Lawesson et al. 2000; Parry & Perkins 2002; Gudjonsson 2009). An increasing need for accurate environmental data has contributed to continuous development of mapping technologies and methods. The national governments allocate large funds to get more or less complete overview of natural resources that has both economic, social, conservation and science benefits. But the mapping process is slow. In Norway, during the last 40 years only 10 % of the country has been mapped, mostly in the mountain regions (Rekdal & Bryn 2010). Each mapping project is closely linked to budget that limits both the mapping process and the recruiting of qualified specialist. The interaction of these, and other factors, causes a need to finding new methods that can accelerate the implementation and increase the efficiency of vegetation mapping. This study explores if one of these new methods that are capable of predicting the spatial distribution of vegetation types using environmental predictors. If the testing of this method ends up with acceptable results, then it can be used as a new tool in resource management and significantly reduce the time and costs during mapping.

The term "vegetation mapping" can be interpreted in different ways, and there is no worldwide consensus (Rekdal & Bryn 2010). The classification of vegetation types was historically closely related to the discipline of phytosociology (Braun-Blanquet 1965; Küchler & Zonneveld 1988). Worldwide, there are many classification systems. In Norway, the most known systems are “Vegetasjonstyper i Norge” (Fremstad 1997) and “Veiledning i vegetasjonskartlegging. M 1:20000-50 000” (Rekdal & Larsson 2005).

In Norway, a growing demand for reliable information about land cover and land resources led to the creation of new national survey system AR18×18 (the Norwegian area frame survey of land resources). This system also constitutes a baseline for studying changes in outfield land resources and a framework for a national land resource accounting system for the outfields (Strand & Rekdal 2010). The AR18×18 survey system is methodologically linked to the European Lucas survey (Land use/cover agricultural survey) carried out in the EU countries by Eurostat (Eurostat 2003). The Lucas system is made up by a grid of 18×18 km mesh size (later condensed to 2×2km) that covers whole Europe and consists of points (the sampling units) located in intersections. These points are the center of a Primary Statistical Units (PSU) of 1500 × 600 meters (0.9 km^2). And in each PSU there are ten sample points. But in Norway, this system was modified, and instead of sample points it was added a detailed land cover of whole PSU at intermediate scale 1:20.000 (Strand 2013; Strand & Rekdal 2010). The Norwegian Forest and Landscape Institute is primarily responsible for conducting the AR18×18 survey in Norway. First this system was tested in the mountains of Hedmark district during the summer season of 2004 and then carried out each year (Strand & Rekdal 2005). The completion of the survey is expected in 2015 depending on available resources (Strand & Rekdal 2010). Despite all advantages, the AR18×18 system does not provide a full overview of the land resources and continuous land cover in the country. But it can give good opportunities to test accelerated mapping processes using geostatistical methods (extrapolation).

The increased use of Geographical information systems (GIS) and new statistical techniques in analyzing geodata has led to a rapid development of predictive distribution modeling (DM) in ecology. A central issue in DM is always related to the analysis of species–environment relationship (Guisan & Zimmermann 2000). The development of DM has also advanced in conjunction with the development of remote sensing-based vegetation mapping (Franklin 1995).

Franklin (1995) used the term predictive vegetation mapping, which has the same meaning as DM, and describes this term as predicting the vegetation composition across a landscape from mapped environmental variables. Usually, common vegetation types are correlated with a large range of environmental variables. This can be challenging, as it can be hard to find specific criteria for the distribution of the vegetation types and difficult to model. But rare vegetation types are often correlated with more narrow variables, which can make them easier to model (Halvorsen 2012a), if you have the variables available in GIS-formats. The knowledge about gradient analysis and ecological niche theory formed the basis of predictive vegetation mapping (Austin 2007; Franklin 1995). The successful use of DM for predictive vegetation mapping, has led these methods to become a very important tool for resource conservation related to e.g. effects of global environmental changes on species distribution and understanding the realized niche of species (Graham et al. 2004; Thomas et al. 2004; Palmer and Van Staden, 1992; Philips & Dudík 2008). The distribution modeling has been used for totally different topics such as the spread of invasive species (Thuiller et al. 2005); biodiversity conservation (Haines-Young 1991), spatial patterns of species diversity (Graham et al. 2006), ecological restoration (Martinez-Taberner et al. 1992), and the potential for expansion of forest following land-use change in Norway (Bryn et al. 2013). DM has also been applied to model land-cover types (Dobrowski et al. 2008) and different species assemblages such as vegetation types (Cawsey et al. 2002; Ferrier et al. 2002; Hemsing & Bryn 2012; Weber 2011).

DM is mostly based on presence observations (presence-only data or PO data) of species occurrences (Stokland et al. 2011). The species occurrence data with good precision is becoming more and more available from different sources like atlases, books, journals, database of museums, as well as digitalized internet sources the GBIF (Phillips & Dudík 2008; Ramsen et al. 2011). At the same time accessible digital maps of environmental variables with high resolution used in DM as predictors, are becoming more common (Bakkestuen et al. 2008; Elith et al. 2006; Crawford & Hoagland 2009; Franklin 2009). The mapping of these variables should be easier than the vegetation mapping itself, in order for DM to be a practical or informative exercise (Franklin 1995). The nature of environmental variables allows to split them into two groups: direct and indirect (Austin 2002). Direct variables influence plants physiologically (temperature, soil, geology and solar radiation). Indirect variables have no directly impact on plants, but by influencing direct variables, they can limit the distribution of species on large geographical scales (latitude and longitude). Using different primary environmental variables it is possible to generate several surrogate variables such as the slope and flow direction extracted from a digital elevation model (DEM). Recently, there has been an increasing use of satellite images in DM (for example Sillero et al. 2012; Stokland et al. 2008). These can be used to obtain several types of predictive variables such as the Normalized difference vegetation index (Weier & Herring 2000) and infrared color bands. The accessibility of PO data and continuous digital environmental data help to provide a rich basis for DM and give possibility for many researchers to use DM for their varying needs.

There are several types of DM with different types of statistical analysis and methods for evaluation of models. Some of these are Expert-based manual modelling (Moravec 1998; Hemsing & Bryn 2012), regression methods such as generalized additive models and generalized linear models (Elith et al. 2006), Rule-based envelope modelling (Bryn 2008) and Statistical predictive modelling (Phillips et al. 2006; Phillips & Dudík 2008). Some of these methods use presence-only data, others also include absences. A special feature of the new methods is their ability to fit more complex models from small datasets and prevent model complexity using mechanisms such as "regularization" (Philips & Dudík 2008).

This master thesis is a continuation and further development of a project related to the possibility of using Distribution modelling (DM) for vegetation mapping (Hemsing 2010; Ullerud 2013). Hemsing (2010) found that a statistical predictive GIS modelling method (MaxEnt) is good to use for the prediction of distribution of vegetation types, and this method was explored further and evaluated for many aspects in Hemsing & Bryn (2012, table 6). Ullerud (2013) studied the possibility to predict distribution of vegetation types in neighboring areas that have no presence data, i.e. spatial transferability. Ullerud (2013) showed that it was possible to extrapolate the DM, but also that the modeling performance varied between different vegetation types. In the modeling Ullerud used presence data from only one sample unit (c. 4 km^2). In this study, instead of using only one sample unit, a grid that consist of many sample units in different densities was tested and the effect of increasing sample unit density on the prediction probability of the whole study area was analyzed.

The main objective of this study is to explore the effect of increasing the density of the PSU with presence-only data on the ability to predict the distribution of three common and three rare vegetation types. The second objective is to answer the following questions:

a) Which density of the PSU grid net is most suitable to predict the distribution of selected vegetation types with the aim of decreasing time and costs for mapping?

b) Which vegetation types can be predicted with high reliability using DM by MaxEnt?

c) Which environmental variables are the most important predictors for the chosen vegetation types?

2. Materials

2.1 Study area

2.1.1 Location of the study area

The study area is located in Venabygdsfjellet at Gudbrandsdalen, in the northwestern part of Ringebu municipality, in the eastern part of Oppland district, Eastern Norway (Figure 1). The center of the study area is Venabu (WGS 1984 UTM-zone 32 Ø555849 N6829297). In total, the study area constitutes 161 km^2. To the west, the study area share borders with Sør-Fron municipality (№ 0519), Oppland district. In the north it shares boarders with Stor-Elvdal municipality (№ 0430), Hedmark district. The southern boundary of the study area goes from Venabygd along road RV27 and Jønnhaltveien to Jønnhalt. The eastern boundary goes from Jønnhalt up along the river Døra, brook Gråbekken to Mykjørrtjønnet and again along road RV27 to Hedmark district. The elevation of the study area is from 330 to 1365 m a.s.l.

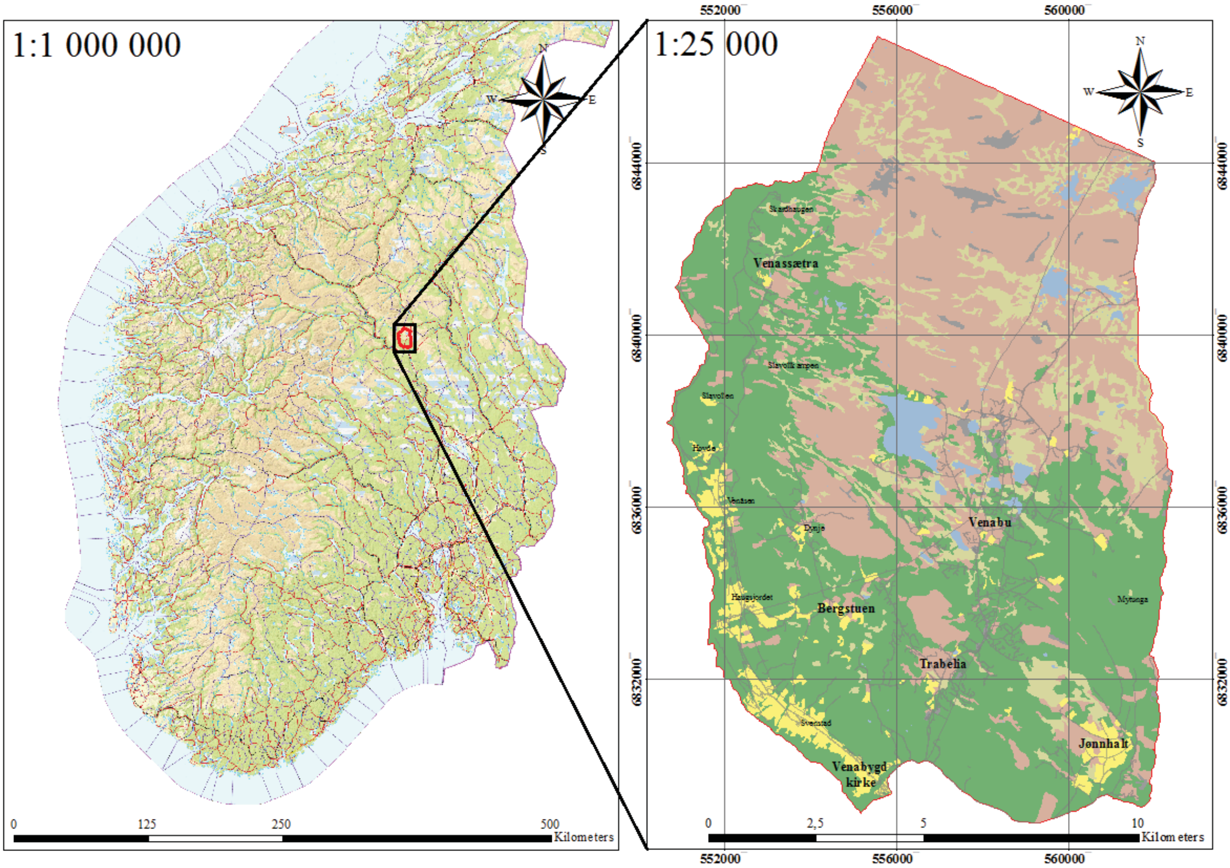


Figure 1. Maps showing the location of study area (red line) in the northwestern part of Ringebu municipality and position in southern central Norway. The maps were created using ArcGIS 10.1 with FKB map data and freely available WMS-service from the Norwegian national geodata coordinator Kartverket (source: www.kartverket.no). Map projection: WGS 84/UTM 32N.

2.1.2 Nature

The study area is characterized by large variations in relief and landform. The area has deep V-shaped valleys with steep slopes surrounded by high mountains in the western and south-eastern parts. The central, northern and northeastern parts have mountains and mountain valleys dominated by mashes and lakes, but without forest. The southern part has flat terrain in some places, more smooth slopes and some mountains. Among the main forms, there are also some small hills and mountain ridges. The study area includes a set of high mountains, e..g. Ramstindan (1334 m a.s.l.), Nørdre Bølhøgda (1365 m a.s.l.), Søndre Bølhøgda (1258 m a.s.l, Svarthammaren (1182), Flakssjølighøgda (1112 m a.s.l.), Dynjefjellet (1147 m a.s.l.), Svartfjellet (1154 m a.s.l.) and Trabelifjellet (1093 m a.s.l.).

The whole study area has many small and several large lakes, narrow rivers and brooks. The largest lakes in Venabygdsfjellet are Flaksjøen (905 m a.s.l.), Bølvatnet (1006 m a.s.l.) and Muvatnet (1052 m a.s.l.) in the northern part. Most of the brooks and small rivers flow to the two large rivers Frya and Nordåa and then flow further down to the great river Lågen in the U-valley at Ringebu and Frya cities. In the V-valleys with deep ravines and gorges, the river flows hastily and foams. Larger waterfalls are not uncommon in these places. Snow and ice sheets are common in the top of valley slopes. The water comes from melted snow that accumulated during winter and from rainfall in warm seasons.

The study area belongs to the north-boreal, low-, median- and high-alpine and south-arctic vegetation zones (Moen 1999). The vegetation in Venabygdsfjellet has a clear zonation. The tree line is around 1050 m a.s.l.. and consist of mountain birch forests that dominates in northwestern part (Bryn & Rekdal 2002). The mountain birch forest grows on the top of valleys slopes, low-lying plateaus and close to mountains. In the low-lying parts the forests, both spruce and pine might occur. The study area also has more grazing influenced birch forests or meadow birch forests influenced by human activity (Puschmann 2005). From around 950 m a.s.l. it is more coniferous trees, and further down spruce forests dominates (Bryn & Murvold 2003). The plunging and slope terrains into the Frya-valley in the west and into Ringebu and Frya cities in the south have birch forest at the top and coniferous forests down to valley bottoms. Pine is common in dry gravel or on scanty and often nutrient-poor rock types and in small quantities in the southern and western part of study area. High and especially slim spruces are a character trait in some places (Puschmann 2005). Along rivers, around cultivated lands, and in parts of properties and around single-homestead deciduous forests dominates (Puschmann 2005). At the bottom of valley gorges, there are elements of alder forests (Bryn & Rekdal 2002). Above the tree line, there is treeless vegetation on or close to mountains. Northern and northeastern parts of the study area are dominated by alpine heaths, especially lichen and dwarf shrub heath around mountain tops, and wetlands in flat areas of mountain plateaus, especially fen and bog. There are also large agricultural areas at Bergstulen, Jønnhalt and between Venabygd and Slavolen along the Frya-valley. At Flaksjøen, Venabu, Trabelia, Bergstulen, Dynje and Jønnhalt, there are also cultivated lands and grazing meadows. In addition, there are several mountain farms spread in the whole study area (Bryn & Rekdal 2002). In older pasturelands, juniper is common and forms cultivated land in some places (Puschmann 2005). The study area was mapped by the Norwegian Institute land Inventory in 2001 (Bryn & Rekdal 2002).

2.1.3 Geology

According to the bedrock map from National Bedrock database (NGU 2014), the study area is dominated by grey sandstone, especially grey meta-sandstone (66 %) and dark grey bad sorted sandstone (22 %). The grey meat-sandstone is spread in the whole study area, especially in the southern part and with a strip in the northern part. The dark grey poorly sorted sandstone is found with two wide strips in the northern part (Appendix 1). Both bedrock types are the sedimentary thrust fault from late Precambrian (Siedlicka et al. 1987). There is also gneiss with a narrow strip from the west near Slavollen to Jønnhalt which covers the area between Trabelia, Bergstuen, Venabu and Jønnhalt, and constitutes 8 % of the study area. In addition, other bedrock types like dolomite, phyllite, granite, conglomerate, quartzite, black slate, schist, meta-gabbro and light sandstone are in small quantities in the west, near Trabelia and between the dominating bedrocks. Totally, they cover ca 3 % of the study area.

The mountain areas are mostly covered by a thin moraine layer, but in the north of Bølhøgda, the layer is thick. On the tops there is exposed bedrocks and boulder fields (Sollid og Trollvik 1991). The mountain areas between lake Flaksjøen and mountain Nødre Bølhøgda have soils with depth lower than 30 cm. In the forest areas, the moraine cover is thick. There are also large marsh areas with organic soil at Jønnhalt, around Venabu, Svartåkluftin, Bølvatnet and down to Mysætrin (Bryn & Rekdal 2002). Especially for the area in the west and south for Trabelia and south of Forbundsfjellet, there are occurrences of large rough boulder fields that have extra good drainage and poor water supply for plant growth. These boulder fields were deposited by glacial rivers at last Ice age (Bergersen 1993).

2.1.4 Climate

The study area is located in a transition zone between a continental and an oceanic climate (Moen 1999). The growing season length constitutes 150-170 days with mean temperature $\geq 5°$ C. The annual mean temperature is -0.28°C. The average temperature for coldest month January is – 9.7°C and warmest month July +10.4°C. The annual precipitation is 700-1500 mm. Number of days with snow cover is 175-225, with more than half of the ground covered with snow (Moen 1999). According to Norwegian Meteorological Institute, that has a meteorological station (930 m a.s.l.., established in 1980) at Venabu, the annual mean temperature in 2013 was 0.15°C, the coldest month was January with average of 11.3°C and the warmest July with 12.2°C (Figure 2). The annual mean wind speed was 2.63 m/s with strongest indicator 15.3 m/s in 27. June. The annual mean precipitation was 68.72 mm with greatest total value of monthly precipitation 163 mm in June and lowest 4.9 mm in March.

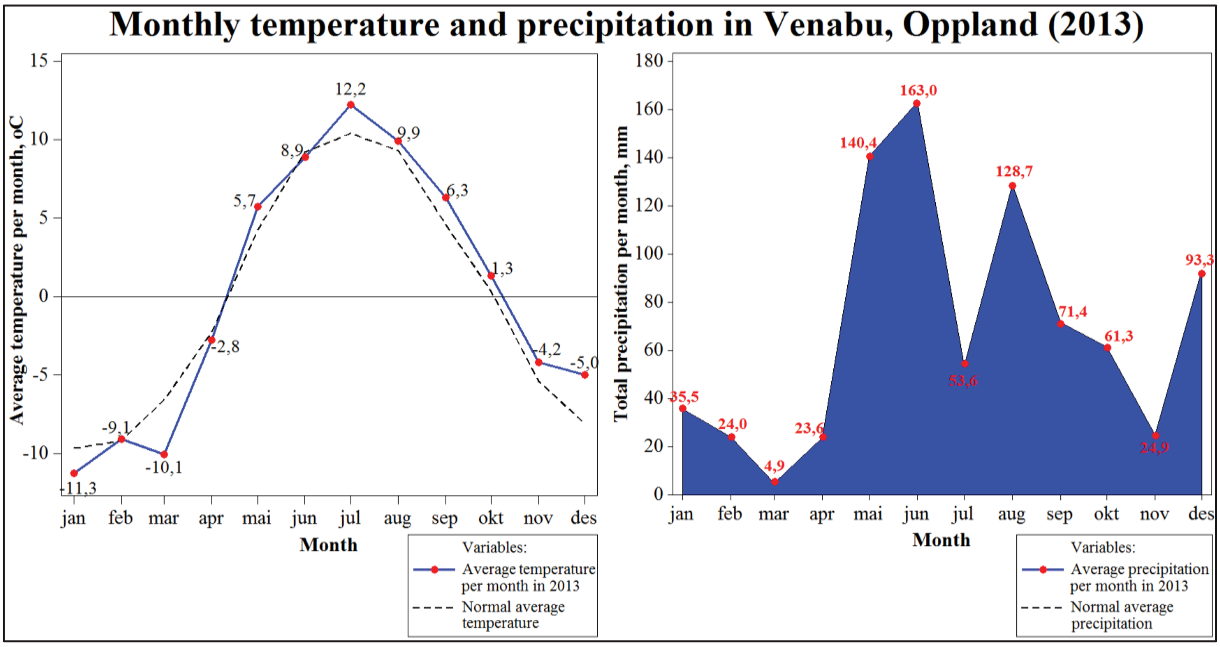


Figure 2. Monthly mean normal and measured temperature and precipitation at Venabu in 2013. Data is taken from Norwegian Meteorological Institute (source: www.met.no)

2.1.5 Cultural influence

The main features of human activities in the study area are farming villages with timber houses and old cultivated lands both near farms and in outfields (Puschmann 2005). From 16th century due to a rapidly increasing population in Ringebu municipality, there was a marked increasing activity of cotter farms in the study area that spread out in the outfields and cultivated new land. Number of farms within study area varied greatly during the last three hundred years. In 1723 there was registered only 9 cotter farms in Venabygd, but in the period 1851-1930 the number increased to 70-80 cotter farms (Hovdhaugen 1988). During 20th century, the number decreased to around 45 cotter farms in 1942, 5 summer dairy farms in 1974 to only 1 active farm in 2006 (Bryn 2006). Therefore, much of the previously cultivated lands are not used anymore, and is becoming reforested by birch. More typically is the spread of farming villages with single farms or small hamlets in between. In some places, large village communities and hamlets are creating greater associated farmlands.

Meadows and pastures dominate in land use areas and the livestock is large, especially in upper mountain areas where cultivated lands are small due to soil specificity (Puschmann 2005). Large areas of cultivated land are generally located near easily accessible places along roads and greater settlements, for example areas along the road Venabygdsveien from Venabygd church to Hovde, at Bergstuen, Dynje, Trabelia and Jønnhalt (Appendix 2).

There are many cottages that are spread in the whole study area and in some places make up cottage fields, especially in areas close to mountains, lakes and on forest covered slopes. The large cottage fields are located in areas around Venåssætra, Dynje, Bergstuen, Trabelia, Friskevarpet and Jønnhalt. Cottage building in outfields is a new trend in Norway and in specially in mountain area (Taugbøl 2002).

In the study area there are two major settlements that represent the current life of region. One of this two is Venabygd. This is a traditional village settlement located on the top of the Frya-valley with traditional agriculture based on husbandry. Also the use of outfield resources like logging, hunting, fishing, the collection of lichens and other types of outfield fodder, and outfield scything and grazing, play an important role among farmers (Almås et al. 2004). The other settlement is Venabu. This is the tourist settlement related to nature experiences and variety of activities like guided ski and snow shoeing tours, dog sleigh rides, biking, swimming, mountain hike etc. The settlement includes many cottages around, the tourist trade, camping and a shopping mall. From Venabu it runs many hiking and skiing trails that cover almost whole study area. A little further north from Venabu, at Flakssjøen there is Venabu mountain hotel with ski resort.

2.2 Vegetation map data

In this study a vegetation map from 2001 was implemented. The mapping was performed by Norwegian Institute for land Inventory (NIJOS) as a result of a project for Ringebu municipality, and in accordance with NIJOS instructions for mapping in scale 1:50 000 (Larsson & Rekdal 1997). The fieldwork was mainly performed in July (Bryn and Rekdal 2002). The vegetation mapping consisted of fieldwork and digitalization of maps. Field registrations were drawn on aerial photos from 1992 (in scale 1:40 000) and then digitized by using an analytical stereo plotter AP 190. The final vegetation database was completed in ArcInfo. The vegetation map was finished in 2002 (Appendix 2). Totally 16 1,46 km^2 (excluding water) was mapped. The largest part of the study area is dominated by forest (45.3 %), especially in the western and southern parts. The other significant part is dominated by alpine vegetation (35.6 %), mostly in the northern and northeastern parts and the center. Wetlands cover 11.2% and are spread in whole study area, both in forest and alpine parts. Agricultural lands cover 4.1 %.

2.3 Environmental variables

In the prediction modeling there were used 12 environmental variables (Table 1, Appendix 4) in the form of raster map with cell size 10x10 m. Totally, each raster map consist of 2 667 249 cells. Some variables were obtained in different resolutions and file formats. Therefore they were transformed to the same resolution and format using ArcGIS's tools. There is a common coordinate system for all raster maps of environmental variables: WGS 1984 UTM Zone 32N. In addition to the right resolution and coordinate system, all raster maps must be within the same boundaries. Otherwise the software used for prediction modeling will not work.

The digital elevation model (DEM) generated from the laser scanning data (Light Detection And Ranging) was used to make the derived variables, namely Basin, Slope, flow accumulation, Flow direction, Curvatures, the TWI. and the NDVI. The satellite image in different bands was used to generate the Normalized Difference Vegetation Index (NDVI). Also there were used three maps that show the distribution of sediments, soil and groundwater in the study area.

Table 1. Overview of environmental variables that were used in the prediction modeling. All variables were transformed to a raster map with the same resolution (10×10), within boundaries and inserted in the same coordinate system before modeling. All preparations were done in ArcGIS.

Environmental variables	**Generated from**	**Original resolution**	**Transformation to 10x10 m**
1. DEM (Digital Elevation model)	LiDAR [1]	10×10	
2. Basin	Flow direction	10×10	
3. Curvatures	DEM	10×10	
4. Flow accumulation	DEM	10×10	
5. Flow direction	DEM	10×10	
6. Groundwater	ND_Løsmasser [3]	continuous vector data	conversion from feature to raster (ArcGIS), with snap to DEM
7. NDVI (the Normalized Difference Vegetation Index)	Satellite image in red (VIS) and infrared (NIR) band. Calculated in the formula: $NDVI = \frac{\text{NIR} - \text{VIS}}{\text{NIR} + \text{VIS}}$	25×25	resample (ArcGIS), with snap to DEM
8. Satellite image: blue, green and red bands			resample (ArcGIS), with snap to DEM
9. Sediments	ND_Løsmasser [2]	continuous vector data	conversion from feature to raster (ArcGIS), with snap to DEM

10. Soil	Berggrunn N50 [2]	continuous vector data	conversion from feature to raster (ArcGIS), with snap to DEM
11. Slope	DEM	10×10	
12. TWI (the Topographic Wetness index)	TWI can quantify the control of local topography on hydrological processes and indicate the spatial distribution of soil moisture and surface saturation: $TWI = \mathrm{Ln}\frac{(\mathrm{fl}_{ac}+1)*10}{\mathrm{Tan}\,(\frac{\mathrm{sl}*1.570796}{90})}$ fl_{ac}- flow accumulation sl – slope 10 – the size of pixels	10×10	

[1] The LiDAR scanning over Venabygdsjellet was done by Statens kartverk in period 2011-2013 (the scanning density 1-5 points per m^2). Generated to DEM (Digital Eleveation Model) in the resolution 10×10 m. Download from Norge Digitalt. "Copyright Norge digitalt"

[2] Beggrunn N50. NGU (Norges geologiske undersøkelser) in scale 1:50.000. The mapping was done be NGU in 1983 and converted to digital form by scanning and vectorization in 2003. Download from NGU net site in shape format (SOSI 4.0). The maps are.

[3] ND_Løsmasser. NGU (Norges geologiske undersøkelser) in scale 1:50.000. The mapping was done by NGU in 1993 and converted to digital form by scanning and vectorisation. Download from NGU net site in shape format (SOSI 4.0).

3. Methods

3.1 Vegetation types

Six vegetation types from three ecosystems (mountain, forest and wetland) were chosen for the prediction modeling (Table 2). In every ecosystem one rare and one common vegetation type was chosen. This choice was based on the present distribution of the vegetation types within the study area. Totally, the summed area of all six selected vegetation types covers 52.7 % of the study area.

Table 2. Overview of the six selected vegetation types and their proportion in the study area.

Type of ecosystem	Vegetation types	Occurrence	Proportion of study area, %
Alpine ecosystem	2e – dwarf shrub heath	common	21.3
	3b – tall forb meadow	rare	1.5
Forest ecosystem	4b – bilberry birch forest	common	22.9
	8d – rich swamp forest	rare	0.4
Wetland ecosystem	9c – fen	common	6.1
	9d – mud- bottom fens and bogs	rare	0.5

3.2 Sample units (PSU) and point grid

For prediction modeling, occurrence information was only gathered for the selected vegetation types within the boundaries of the sample units. The sample units used in data collection is equivalent to a Primary Statistical Unit (PSU) of the AR18×18 survey system of 1500×600 m size (Figure 4). The PSU have been stratified according to different densities. Each PSU was inserted in bottom left corner of a grid mesh. The distance between sample units was measured from and to these corners (Figure 3).

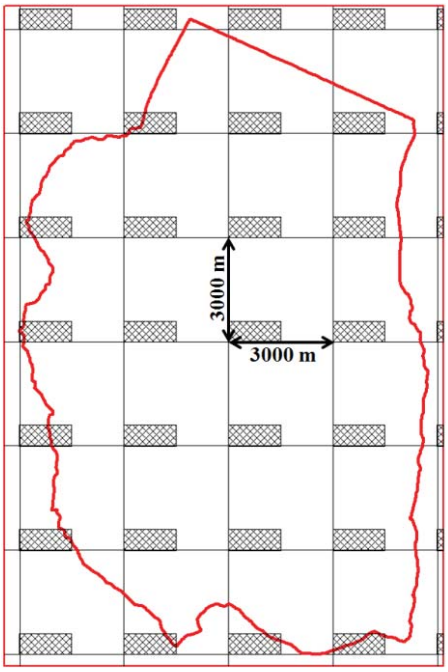
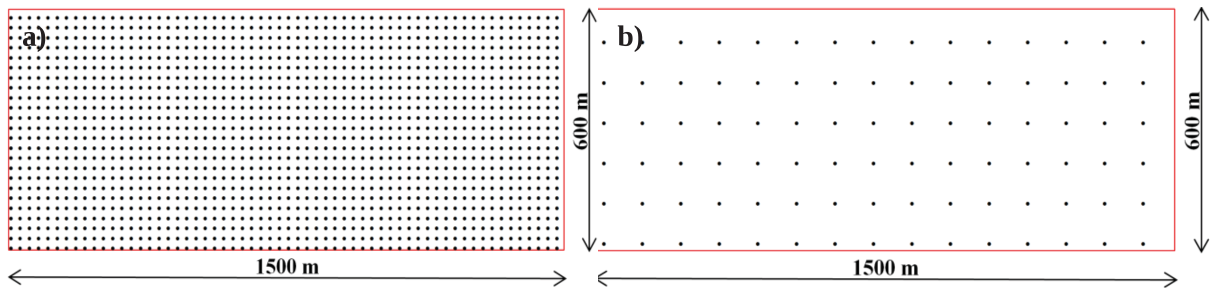


Figure 3. The study area (red line) covered by a 3x3km grid mesh (black line). The distance between sample units (latticed rectangles) is taken from and to bottom left corner of each unit.

In the MaxEnt software (see the next paragraph), that was used in the prediction modeling, the information about presences of selected vegetation types is taken in the form of point grid and presented as PO data (presence-only data). Due to significant difference in the occurrence of selected vegetation types, rare and common vegetation types were tested using different presence point density. As a result of these tests, rare vegetation types was tested using a point grid distance of 25 meters, whereas 100 meters point grid distance was used for common vegetation types (Figure 4).



Figur 4. The sample unit with to two types of point grid: a) with 25m distance for rare vegetation types; b) with 100m distance for common vegetation types. The sample unit that was used in prediction modeling is originally taken as a Primary Statistical Unit (PSU) in AR18×18 system. The size of each unit is 1500×600 m.

On the vegetation map the selected vegetation types is shown sometimes as mosaics and with additional signs. The mosaics consist of two or more vegetation types that are spatially mixed with each other, so that they can't be separated into different polygons. The additional signs were used to show important characteristics of the vegetation cover that are not used for the general description of the vegetation types. Both the selected vegetation types represented as the secondary vegetation types in mosaics and additional signs were not taken into consideration and were excluded during the creation of PO data (Table 3). This exclusion is seen as a necessity to improve the models to recognize selected vegetation types in their specific ecological ranges and increase the predictive performance of models. The creation of both sample units and point grid, and all other processes related to geographical and statistical analyses were done in ArcGIS software (version 10.1).

Table 3. The number of training points (PO data) generated for prediction modeling in each selected vegetation type within sample units, their kind of mosaic and additional signs that were and not were involved in modeling. Training points were created in ArcGIS (version 10.1). Description of codes an additional signs is given in Appendix 3.

Vegetation type	Density of sample units	Number of training points from sample units taken into prediction modeling	Mosaics and signs included in presences	Additional signs
2e – dwarf shrub heath	3×3 4,5×4,5 6×6 7,5×7,5 9×9	374 115 59 91 52	2e, 2e/2c, 2e/3b, 2e/9c	g, j, H, k, n, s, v, ɔ, o), *, ⟠, +
3b – tall forb meadow	3×3 4,5×4,5 6×6 7,5×7,5 9×9	355 53 48 21 not found	3b, 3b/2e, 3b/9c	g, j, H, k, s, ɔ, *
4b – bilberry birch forest	3×3 4,5×4,5 6×6 7,5×7,5 9×9	370 164 30 34 1	4b, 4b/2e, 4b/4a, 4b/4c, 4b/6b, 4b/8d, 4b/9c,	g, v, o), *,⟠, +,]

8d – rich swamp forest	3×3 4,5×4,5 6×6 7,5×7,5 9×9	331 45 not found not found 25	8d, 8d/7b, 8d /9c	k, o)
9c – fen	3×3 4,5×4,5 6×6 7,5×7,5 9×9	132 26 18 19 12	9c, 9c/2e, 9c/3b, 9c/9a, 9c/9d	ɔ, o), k, s, g,
9d – mud-bottom fens and bogs	3×3 4,5×4,5 6×6 7,5×7,5 9×9	381 264 not found 109 264	9d, 9d/9a, 9d/9c	none

3.3 Distribution modeling

Today there are several methods for DM of vegetation types, but the best method was recognized a statistical predictive GIS-modelling method (Hemsing & Bryn 2012), a method for presence-only DM. This method is based on maximum entropy modeling, often described as a machine learning method, which are trying out various interactions between environmental factors (Phillips et al. 2006; Phillips & Dudík 2008), and can also be explained as a maximum likelihood method (Halvorsen 2012a).

In this study the MaxEnt software (version 3.3.3k) was used to prepare statistical predictive models for the potential distribution of vegetation types (Phillips et al. 2004). The main idea of the modeling is to estimate an unknown probability distribution of, in our case, vegetation types in relation to a set of restrictions. The restrictions put in MaxEnt are that the expected value (the true mean) related to each environmental variables should be the same as the observed mean (Stokland et al. 2008). It is more about statistical analysis of combinations and interactions of environmental variables in the presence-cells and finding of locations where the target might be present (Elith et al. 2011; Phillips & Dudik 2008).

The models in MaxEnt were created in the form of rasterized frame-area for training and used environmental variables for this frame-area for projection. Each cell marks MaxEnt as an observation unit. There are two types of observation units: presence and absence. By using extrapolation the frame-area generate map representations of model predictions. The prediction results from MaxEnt are given as relative predicted probabilities of presence (RPPP) because models are based on PO data and the prevalence of the modelled target is not known. It is known nothing about eventual presence or absence in the uninformed background observation units (Phillips et al. 2006; Ward et al. 2009). The term “relative” means that model predictions can be compared among grid cells, but that their absolute values cannot be used for interpretation in terms of probabilities of presence of the modelled target (Ferrier et al. 2002).

In order to translate RPPP to predicted probability of presence (PPP), the modelled predictions were evaluated with independent data. This independent data was generated from the vegetation map where there were spread random points in the frame-area for projection and attaching presence/absence information to each point. Evaluation of models results is based on the MaxEnt output using the following parameters:

1) ROC curves and AUC-value

Evaluation of results from MaxEnt was done by a threshold-independent receiver operating characteristic (ROC) analysis with ROC curve and AUC-values. The ROC curve evaluates a models usefulness to predict the relative distribution probability of species (Elith et al. 2006). The curve is obtained by "joining the dots" (Phillips et al. 2005) and plotting the species true positive rate on the y-axis and the false positive rate on the x-axis for all possible thresholds (Phillips et al. 2006). In other words, the curve shows how the sensitivity and specificity varies as a function of the threshold.

The area under the ROC curve is AUC, which measures the quality of a ranking of sites or the models relative predictive ability (Fielding and Bell 1997; Franklin 2009; Halvorsen 2012a; Pearce & Ferrier 2000). In MaxEnt AUC-values are not based on a normal ROC curve, but on a presence-versus-random ROC curve (Phillips et al. 2006). The AUC is the probability to differentiate between presences and pseudo-absences, and that a randomly chosen presence site will be ranked above a randomly chosen absence site (Halvorsen 2012a; Stokland et al. 2011). It is important to note here that AUC-value become higher for predictive object (species) that have narrow ranges of environmental parameters (Phillips et al. 2006). AUC-value ranges from 0 up to 1. The closer to 1 the AUC-values are, the greater the model's predictive ability is, whereas AUC-value 0.5 is equal to a random model (Pearce & Ferrier 2000). Models with values above 0.75 are considered potentially useful (Elith 2002). In this study results should be evaluated by the classifying scale shown in Table 4.

Table 4. The scale for classifying of DM's result from MaxEnt based on the AUC-value.

Classes	Worthless	Poor	Fair	Good	Excellent
AUC-value	< 0.60	0.61-0.70	0.71-0.80	0.81-0.90	> 0.91

2) Map representation of the prediction model

In addition to the statistical results, the MaxEnt exports a raster representation of the model that takes the form of a map (Appendix 5). This map shows the predicted distribution of the modeled vegetation types. Variation of predicted probability of presences (RPPP) conditions is shown in different colors. Red color is indicating high probability of suitable conditions for the certain vegetation type, green indicating conditions typical of those where certain vegetation type is found, and lighter shades of blue indicating low predicted probability of suitable conditions (Phillips et al. 2006). White dots on a map show the presence locations used for training and while violet dots show test locations.

3) Response curves

Response curves show how each environmental variable affects the MaxEnt prediction and which variable becomes a good predictor (Phillips et al. 2006). Each curve presents a different model. There are two types of response curves: with marginal and single effect. Marginal effect means that variation of each environmental variable will follow to changes in the logistic prediction, while all other variables are kept constant. Response curves with single effect show the response to only one environmental variable.

4) Analysis of variable contributions

After response curves there comes a table that shows estimates of relative contributions and permutation importance of the environmental variables to the MaxEnt model. By modifying the coefficient for single feature each step of modeling algorithm increases the gain. The MaxEnt relates the increase in the gain to environmental variables that feature depends on. And at the end of training process it become converted to percentages (Phillips et al. 2005).

5) "Jack-Knife test"

"Jack-Knife test" was used for to evaluate variables importance and contribution to the model in MaxEnt (Phillips & Dudik 2008; Halvorsen 2012a). Results of this test come out as a graphic that shows the gain of each environmental variable in isolation and point out certain variables, which appear to have the most useful information by itself and that isn't present in the other variables. Environmental variables with low contribution (less than 0.005 to the AUC-value) to the model were excluded and not used in the ultimate model tests, following Stokland et al. (2011).

Before the final model testing, training tests were run to determine the right settings for each vegetation type. The ascertainment of the right settings was based on the evaluation of the same parameters listed above and comparison output raster map with existing vegetation map.

The logistic output format was used in DM. The reason for this choice is that it was easier to analyze output results and conduct further statistical comparison with real distribution of selected vegetation types when the probability value of presence is represented in scale from 0 to 1. Actually this value shows the percentage of probability value to find certain vegetation type in a given place in the modelled reality (Phillips et al. 2005).

Based on training experiments related to improve prediction ability there was chosen three types of features: linear, quadratic and product. The linear feature is equal to continuous environmental variables and ensures that the mean value of environmental variables at where the vegetation type is predicted to occur approximately matches the mean value where it's observed. The threshold feature makes a continuous predictor binary derived by thresholding environmental variables and gives value 1 above the threshold and 0 below. The quadratic feature is the square of the linear environmental variables. This feature constrains the variance in environmental variables where the vegetation type is predicted to occur to match observation. The product feature is equal products of pairs of continuous environmental variables. In other words, this feature constrains the covariance of environmental variables with other predictors and is equivalent to interaction terms in regression. The hinge feature is like threshold feature, except that a linear function is used (Phillips & Dudik 2008; Halvorsen 2012b; Merow et al. 2013).

Threshold- and hinge features were not activated because they often led to overfitted models. Changing of parameter "regularization multiplier" (RM) under threshold and hinge features didn't lead to less overfitted models. This parameter is used to avoid overfitting in MaxEnt. RM affects how focused or closely-fitted the output distribution is (Phillips et al. 2005). The default value in MaxEnt's settings is 1.0. A smaller value than 1.0 will result in more localized output distribution that is a closer fit to the given PO data and the model doesn't generalize well to independent test data. A larger value than 1.0 will result in more spread out distribution, less localized prediction. In addition, the potential for overfitting increases as the model complexity increases.

Other functions and parameters that are available in MaxEnt were not used. For those functions and parameters, the default setting was used.

3.4 Comparison of the predicted and real distribution

The analysis of the MaxEnt results consists of a statistical comparison of the output raster map with the ground truth given by the implemented vegetation map. In other words, it was carried out comparison of the predicted distribution with the real distribution of selected vegetation types (overlay).

During the preparation of data first it was created a point grid with 10 meters distance converted from raster map from MaxEnt's output using conversion tool in ArcGIS. Then this point grid was cut out in boundaries of the study area. In this way it was covered the whole predicted distribution map and incorporated all probability values (RPPP) of each one cell. This conversion process was performed separately for each model. Using the join-function in ArcGIS, the information about the real distribution from the vegetation map was inserted into the same point grid. As result it was obtained one point grid that contains information about both predicted probability of presences (RPPP) of modelled vegetation type and real presence (PO data) in a given place. This information was exported as a table that was further analyzed statistically.

In the statistically analysis all points were classified to 5 classes (Table 5) according to probability value. Then all points were summed with the PO data within each class. A successful prediction model was considered as a model with such distribution that has an increasing number of PO data form low to high probability value, in other words has high probability values geographically located within the boundaries of real distribution of the modeled vegetation type. While the number of points, that are outside the real distribution, should decrease from low to high value. The probability values located in the cultivated lands and water surfaces were excluded from comparison analysis. The final evaluation of the prediction models is based on results from MaxEnt's evaluation and the comparison analysis.

Table 5. The scale for classifying of probability value from MaxEnt output, which was previously transformed by logistic output format in the scale from 0 to 1.

Classes	Worthless	Low	Middle	Good	Excellent
Probability value	< 0.60	0.61-0.70	0.71-0.80	0.81-0.90	> 0.91

4. Evaluation of the implemented vegetation map

The presence data used for the prediction modeling is taken from the digitalized vegetation map implemented in 2001. Before model testing, there was the need to evaluate the quality of this map. To evaluate the quality of the implemented vegetation map, fieldworks was conducted, and followed by subsequent statistical analyses. The fieldwork was conducted in five locations: Venassætra, Bølvatnet, Flaksjølia, Bergstulen and Jønnhalt (Figure 5). All five locations have differences in landscape, environmental and climatic conditions and were subjectively selected to represent as much variation as possible from the study area. Bølvatnet and Flaksjølia represent mountain environments with elements of alpine heath, meadow communities and wetlands, steep terrain in some places and mostly without a bush layer. But, bush layers occur along small brooks that flow into the lakes.

Jønnhalt includes large wetlands, especially bogs and fen marshes, deciduous forest, alpine heath and meadow communities, farm lands, and the landscape is more flat than the two previous sites. Venassætra and Bergstuen have a lot of deciduous- and spruce forests, alpine meadow communities, and in small quantities also becomes elements of wetlands and pine forests. Bergstulen also includes cultivated land.

The observation points were generated randomly in ArcGIS (function: generate random points) and then transferred to the GPS. Later on, the observation points were joined and intersected with the vegetation map. To find the validation points, there were used the GPS navigator and the detailed topographic paper map.

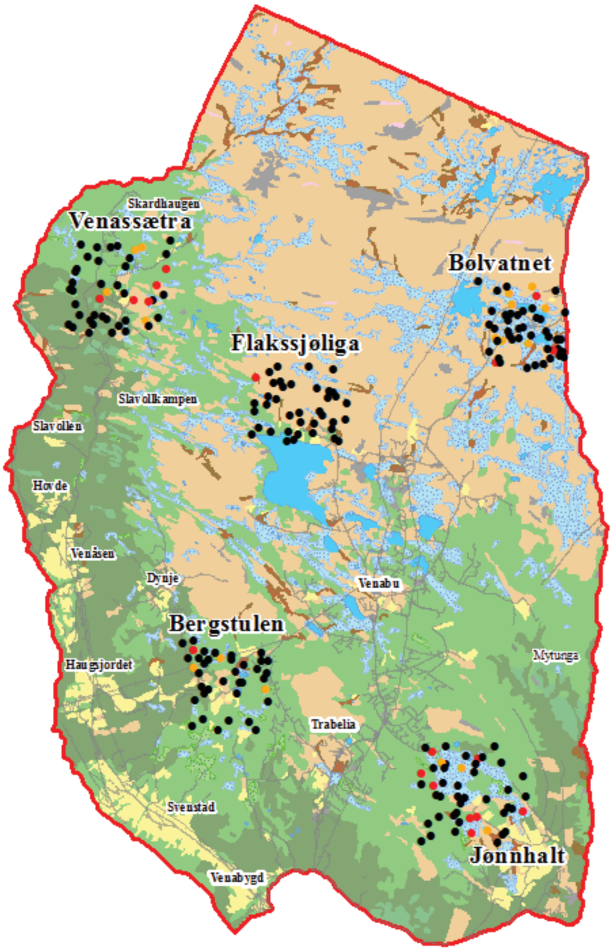


Figure 5. The map shows five locations where it was conducted fieldwork related to the evaluation of quality of old vegetation map. Each observation site is represented in different color: black – correspondence, orange – insignificant error and red – significant error.

Totally, 220 points were observed (Table 6). As a result of the statistical analysis, all observation points were classified into three groups:

1- correspondence:

- observed vegetation type corresponded with the vegetation type in the map
- specified vegetation type was similar to the observed vegetation type, but with different additional symbols

2 - insignificant errors:

- specified vegetation type was observed in less than 10 meters away from its designated area on the map (this distance is given because of the uncertainties in navigation of the GPS and the resolution intended from the mapping of vegetation)
- specified vegetation type is the primary dominant vegetation type in a mosaic (the mix of two or more vegetation types), but with different secondary vegetation types

3 - significant errors:

- vegetation type didn't exist in a given place on the map
- specified vegetation type was observed in more than 10 meters away from its designated area on the map

Table 6. Overview of observed locations and classified errors per location and vegetation type.

Locations	**Number of observations**	**Total number of errors**	**Number of significant errors**	**Number of insignificant errors**	**Number of errors per vegetation type**
Bergstulen	38	4	1	3	1 - alpine vegetation 1 - deciduous forest 2 - spruce forest
Bølvatnet	50	10	3	7	6 - alpine vegetation 4 - wetlands
Flaksjølia	39	1	1		1 - alpine vegetation
Jønnhalt	48	11	8	3	2 - alpine vegetation 5 - deciduous forest 4 - wetlands
Venassætra	45	9	6	3	3 - alpine vegetation 4 - deciduous forest 2 - pine forest
Total	220 (100 %)	35 (15.9 %)	19 (8,6 %)	16 (7,3 %)	13 - alpine vegetation 10 - deciduous forest 2 - spruce forest 2 - pine forest 8 - wetlands

From 220 observation points there were registered 35 errors (15.9 %), where 19 - significant errors and 16 - insignificant errors. The largest numbers of errors were registered in alpine heath communities (13 errors) and deciduous forest there were registered (10 errors). Also it was registered 8 errors in wetlands, 2 errors in spruce forest and 2 errors in pine forest. The greatest number of errors was registered in Jønnhalt (11 errors) and least in Flaksølia (only 1 error). Most of insignificant errors were found in wetlands (9 cases). The number of correspondences is 185 that corresponding to over 84 % of total number of all observation points.

5. Results

5.1 MaxEnt result

In total, the modeling resulted in 26 predictions based on different SPU grid net size (Table 7). Some vegetation types (4) did not have any training points (PO data) in some of the grid densities, and thus did not result in any model output. Of these 26 models only four had AUC-values less than 0.80 (poor models). The other models had AUC-values above 0.80, can be classified as good models by MaxEnt, and had strong relationship between environmental predictors and the modeled vegetation types. Also the prediction modeling has shown that the models of the rare vegetation types had very high AUC- and RPPP-values and therefore mostly were classified as excellent models. The models of the common vegetation types had less both AUC- and RPPP-values.

Table 7 shows that the number of training points decreases with decreasing density of a Primary Statistical Units (PSU). An exception from this overall trend is documented by vegetation type 8d, which has the same number of training points in PSU grid-densities 4.5×4.5 km and 9×9 km. Among all vegetation types the greatest number of training points is in the 3×3 km PSU grid. Vegetation type 4b had only 1 training point in the 9×9 km PSU grid, which led to a useless model. The highest AUC-value was gained for vegetation type 9d in a 7.5×7.5 km PSU grid.

Table 7. Evaluation of MaxEnt's results. This table shows the variation in AUC values as result of varying of number of training points in the model at a constant set of environmental variables (predictors). The most important variables for each vegetation type is provided in the last column.

	Vegetation type	PSU grid density, km	Number of training points in the model	AUC value	Highest RPPP value	Classification of DM's result	Most important environmental variables
Common vegetation types	2e – dwarf shrub heath	3×3	374	0.807	0.76	good	DEM, NDVI, Blue band
		4,5×4,5	115	0.872	0.74	good	
		6×6	59	0.739	0.69	fair	
		7,5×7,5	91	0.845	0.91	good	
		9×9	52	0.961	0.82	excellent	
	4b – bilberry birch forest	3×3	370	0.826	0.84	good	DEM, NDVI, Blue band
		4,5×4,5	164	0.861	0.87	good	
		6×6	30	0.662	0.86	poor	
		7,5×7,5	34	0.890	0.74	good	
		9×9	1	0.500	0.50	worthless	
	9c – fen	3×3	132	0.898	0.95	good	DEM, Slope, Sediments, Red band
		4,5×4,5	26	0.870	0.89	good	
		6×6	18	0.957	0.99	excellent	
		7,5×7,5	19	0.954	0.99	excellent	
		9×9	12	0.814	0.86	good	

Rare vegetation types	3b – tall forb meadow	3×3	355	0.871	0.99	good	DEM, NDVI, Blue band, Slope, TWI
		4,5×4,5	53	0.950	0.91	excellent	
		6×6	48	0.946	0.94	excellent	
		7,5×7,5	21	0.920	0.99	excellent	
		9×9	not found	–	–		
	8d – rich swamp forest	3×3	331	0.955	0.96	excellent	DEM, NDVI, Blue band, Slope
		4,5×4,5	45	0.926	0.98	excellent	
		6×6	not found	–	–		
		7,5×7,5	not found	–	–		
		9×9	25	0.992	0.93	excellent	
	9d – mud-bottom fens and bogs	3×3	381	0.989	0.98	excellent	DEM, Blue band, Sediments, Slope, Groundwater, Soil
		4,5×4,5	264	0.993	0.99	excellent	
		6×6	not found	–	–		
		7,5×7,5	109	0.994	0.99	excellent	
		9×9	264	0.993	0.99	excellent	

The most important environmental predictors vary among the modeled vegetation types, but the most common were altitude (digital elevation model; DEM), Normalized Difference vegetation Index (NDVI), LandSat image (blue band) and Slope. From these predictors, only DEM is important for all models, but contributing in varying degree to the model performance. Among marsh communities (9c, 9d), other environmental predictors, such as sediment, soil and groundwater were included, based on increased model performance.

5.2 Comparison of predicted and true distribution of vegetation types

In the statistical analyses, the data with relative probability values (RPPP) was based on the output raster maps (10×10m resolution) from MaxEnt. These were projected into a point grid that afterwards was clipped within the boundaries of the study area. Totally, it resulted in a point grid with 997 638 points from a 10×10m plot mesh. The proportions of the modeled vegetation types given by these points differs from the real distribution of the vegetation types (Table 8) provided by the original vegetation map. In a point grid the vegetation type 2e includes 303 193 points (30.39 %) in point grid, 4b – includes 159 201 points (15.96 %), 9c – includes 84 869 points (8.51 %), 3b – includes 18 035 points (1.81 %), 8d – includes 932 points (0.09 %) and 9d – includes 6 962 points (0.70 %).

The main idea behind this statistical analysis is to evaluate if the number of points within the true distribution increases with increasing RPPP value, while the number of points outside the true distribution (still with increasing RPPP) decreases. This method represents an independent evaluation of the model performance, by comparing the model output for the different vegetation types with the true distribution. However, such a trend, i.e. an increasing RPPP within the true distribution, is only registered in four out of 26 models (Table 8). These four vegetation types were:

- Vegetation type 2e in a PSU grid density 4.5×4.5 km
- 4b - 3×3 km and 4.5×4.5 km
- 3b - 3×3 km

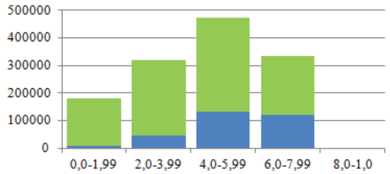
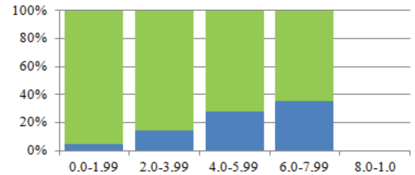
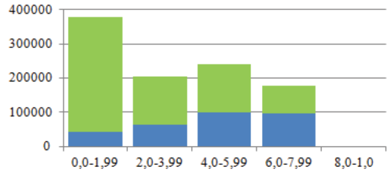
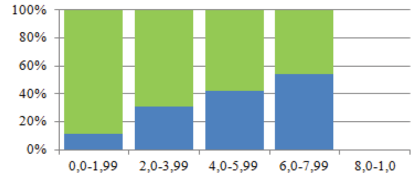
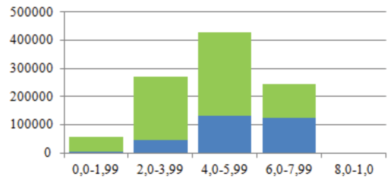
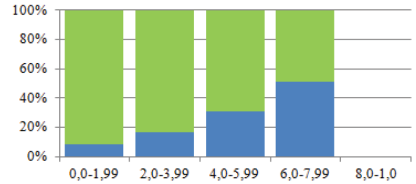
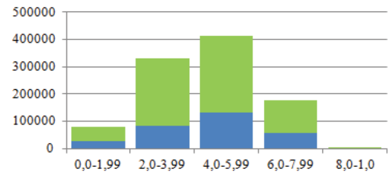
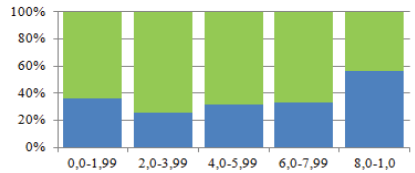
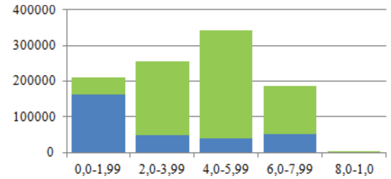
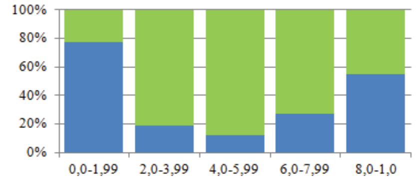
The other models have more varying results, mostly a decreasing number of true points with towards higher RPPP values. Therefore, these models should not be considered as good models regardless of their potentially high AUC-values. In ten of the models the number of RPPP points within the true distribution increases only to the fourth class of RPPP value (0.6-0.8), and then decreases again in the fifth class (0.8-1.0 RPPP) or have no point in this class.

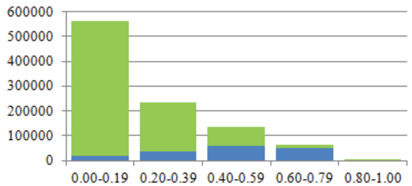
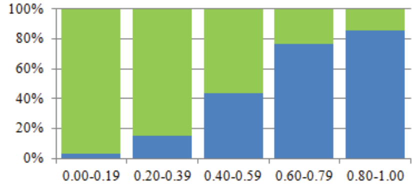
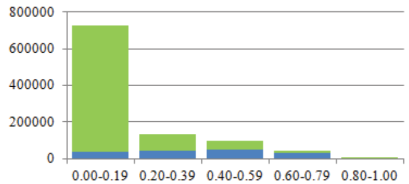
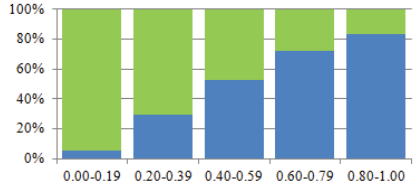
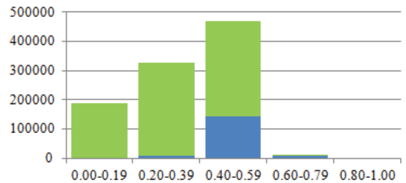
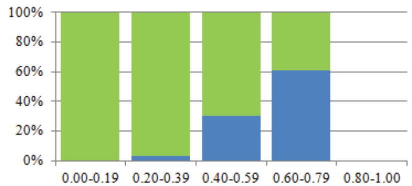
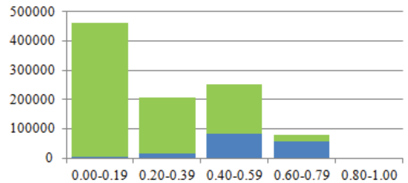
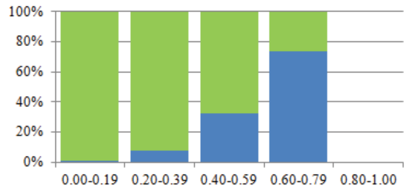
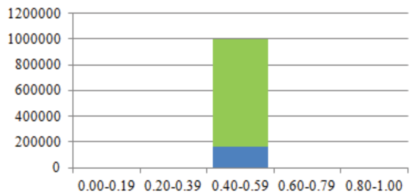
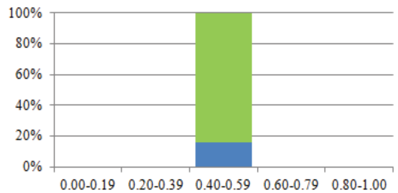
Out of 26 models, no model should be evaluated as better than good. Thus, the analysis including the evaluation data, show that none of the tested models is good enough for a reasonable modelling of vegetation types, following the executed methods with the provided environmental layers. Out of 26, 7 models however, were evaluated as good:

- Vegetation type 2e in a PSU grid densities 3×3 km and 4.5×4.5 km
- 4b - 3×3 km and 4.5×4.5 km
- 9c - 3×3 km and 7.5×7.5 km
- 3b - 3×3 km

These results show that the prediction models perform better with more dense grid meshes of PSUs, than with more scattered grid meshes. The other 19 models have been interpreted as unsuccessful, since few of the points with high RPPP values falls within their true distribution. Interestingly, the vegetation types 8d and 9d got very poor evaluation results, whereas they resulted in the highest AUC-values among all MaxEn models. The overall summary of results is presented in the table in the next page (Table 8).

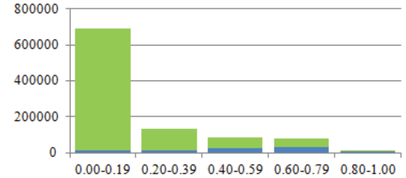
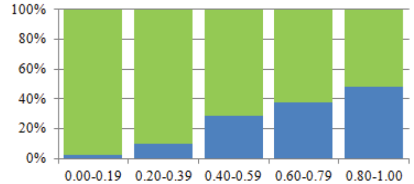
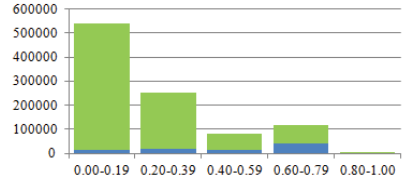
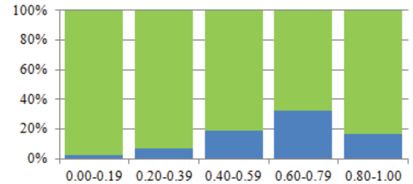
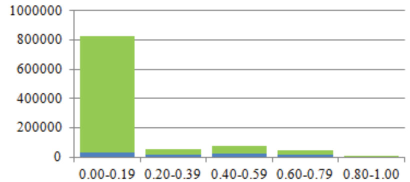
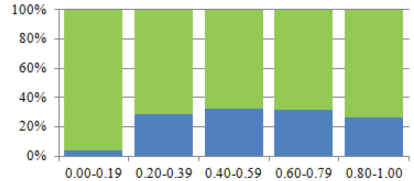
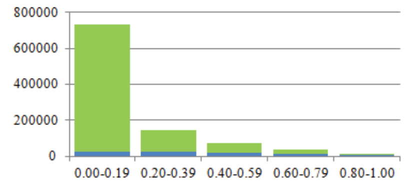
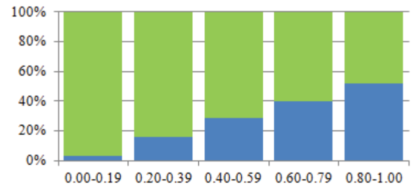
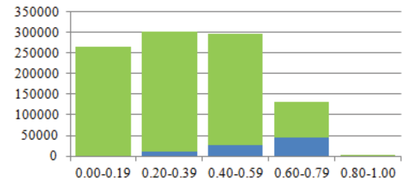
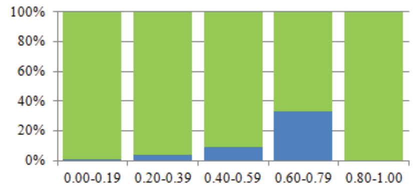
Table 8. The statistical comparative analyses of predicted (RPPP) and true distribution of the modeled vegetation types. The table gives information about the number of cells of each vegetation type in a raster map, the number of cells inside and outside of true distribution in 5-class scale of probability value (RPPP). The last column shows the final interpretation of the MaxEnt models.

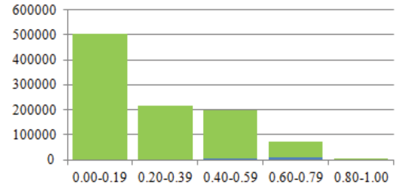
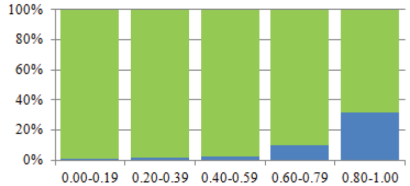
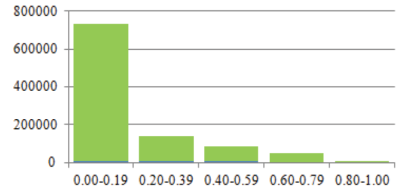
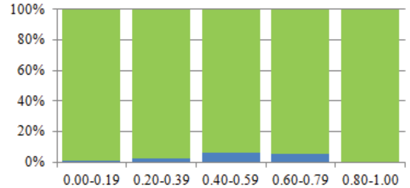
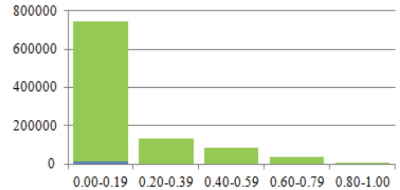
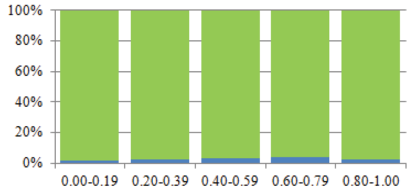
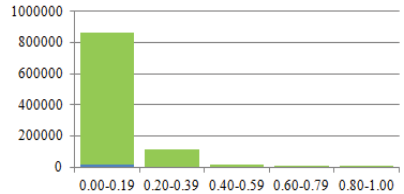
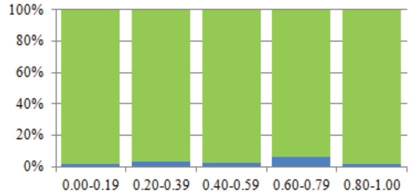
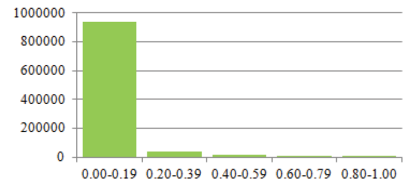
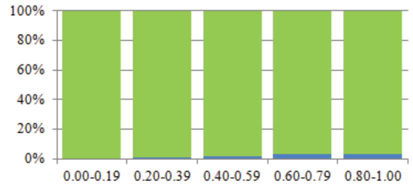
Vegetation type	Density of sample units	AUC-value	Quantitative and percentage relationship between points within and outside for true distribution of modeled vegetation types in 5-class scale of probability value (RPPP): 0.0 - 0.2	0.2 – 0.4	0.4 – 0.6	0.6 – 0.8	0.8 – 1.0	Quantitative distribution of points within and outside of true distribution of modeled vegetation types (outside / within)	Percentage of points within and outside of true distribution of modeled vegetation types (outside / within)	Final evaluation
Common vegetation types: 2e – dwarf shrub heath	3×3	0.807	8017 (0.80%) / 162480 (16.29%)	45192 (4.53%) / 228242 (22.88%)	131619 (13.19%) / 208381 (20.89%)	118265 (11.86%) / 95342 (9.56%)	- / -			good
	4,5×4,5	0.872	43213 (4.33%) / 333476 (33.43%)	63472 (6.36%) / 140420 (14.08%)	100643 (10.09%) / 139861 (14.02%)	95865 (9.61%) / 80688 (8.09%)	- / -			good
	6×6	0.739	4767 (0.48%) / 51176 (5.13%)	44286 (4.44%) / 226218 (22.68%)	130697 (13.10%) / 298108 (29.88%)	123443 (12.37%) / 118943 (11.92%)	- / -			unsuccessful
	7,5×7,5	0.845	28378 (2.84%) / 50740 (5.09%)	83881 (8.41%) / 246034 (24.66%)	130699 (13.10%) / 281004 (28.17%)	58283 (5.84%) / 116667 (11.69%)	1952 (0.20%) / 1502 (0.15%)			unsuccessful
	9×9	0.961	162717 (16.31%) / 48458 (4.86%)	48447 (4.86%) / 208080 (20.86%)	40418 (4.05%) / 302592 (30.33%)	50000 (5.01%) / 135315 (13.56%)	1611 (0.16%) / 1344 (0.13%)			unsuccessful

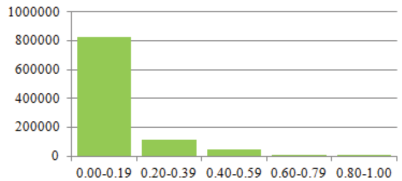
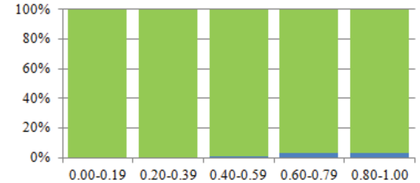
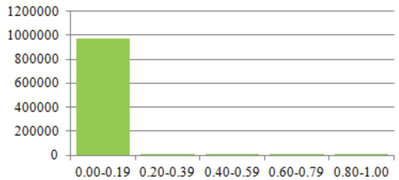
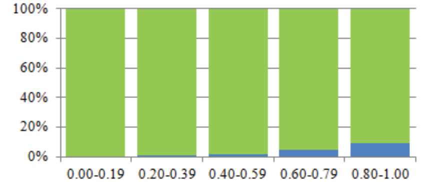
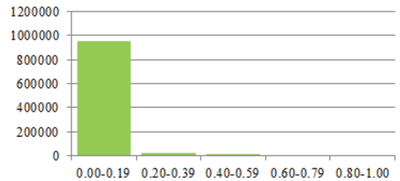
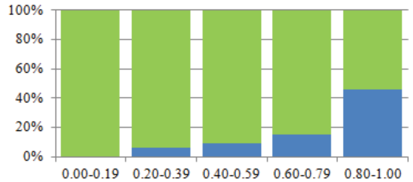
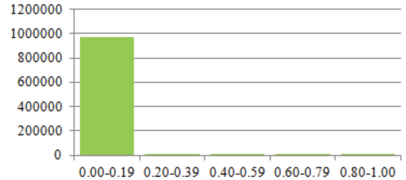
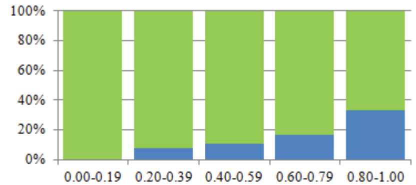
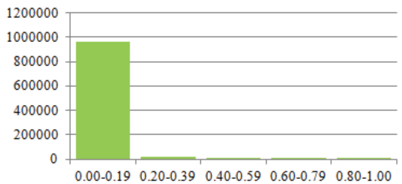
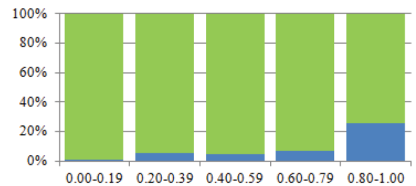
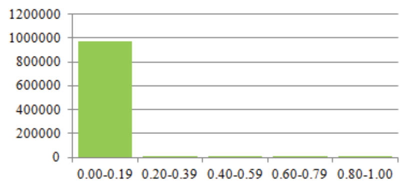
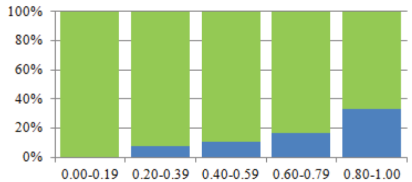
Common vegetation types **4b – bilberry birch forest** 3×3	0.826	16206 (1.62%) / 546523 (54.78%)	36038 (3.61%) / 199732 (20.02%)	59135 (5.93%) / 77786 (7.80%)	47804 (4.79%) / 14393 (1.44%)	18 (0.0018%) / 3 (0.0003%)			good
4,5×4,5	0.861	39029 (3.91%) / 686716 (68.83%)	39571 (3.97%) / 94259 (9.45%)	50667 (5.08%) / 45926 (4.60%)	29800 (2.99%) / 11509 (1.15%)	134 (0.01%) / 27 (0.003%)			good
6×6	0.662	597 (0.06%) / 188505 (18.90%)	9635 (0.97%) / 318382 (31.91%)	141951 (14.23%) / 327070 (32.78%)	7018 (0.70%) / 4480 (0.45%)	- / -			unsuccessful
7,5×7,5	0.890	4060 (0.41%) / 458043 (45.91%)	16004 (1.60%) / 189831 (19.03%)	82091 (8.23%) / 170128 (17.05%)	57046 (5.72%) / 20435 (2.05%)	- / -			unsuccessful
9×9	0.500	- / -	- / -	159201 (15.96%) / 838437 (84.04%)	- / -	- / -			unsuccessful

Common vegetation types

9c – fen

3×3	0.898	14280 (1.43%) / 677354 (67.90%)	12807 (1.28%) / 121020 (12.13%)	24811 (2.49%) / 62039 (6.22%)	28814 (2.89%) / 47796 (4.79%)	4157 (0.42%) / 4560 (0.46%)			good
4,5×4,5	0.870	13058 (1.31%) / 529105 (53.04%)	16495 (1.65%) / 233746 (23.43%)	15918 (1.60%) / 67206 (6.74%)	38950 (3.90%) / 80454 (8.06%)	448 (0.04%) / 2258 (0.23%)			unsuccessful
6×6	0.957	31734 (3.18%) / 793655 (79.55%)	14239 (1.43%) / 36134 (3.62%)	23809 (2.39%) / 49630 (4.97%)	14389 (1.44%) / 31406 (3.15%)	698 (0.07%) / 1944 (0.19%)			unsuccessful
7,5×7,5	0.954	21690 (2.17%) / 712984 (71.47%)	23428 (2.35%) / 122369 (12.27%)	20538 (2.06%) / 50573 (5.07%)	15443 (1.55%) / 23366 (2.34%)	3770 (0.38%) / 3477 (0.35%)			good
9×9	0.814	1352 (0.14%) / 264180 (26.48%)	11885 (1.19%) / 290793 (29.15%)	27611 (2.77%) / 269232 (26.99%)	44021 (4.41%) / 87730 (8.79%)	0 (0.00%) / 834 (0.08%)			unsuccessful

Rare vegetation types	**3b – tall forb meadow**	3×3	0.871	2213 (0.22%) / 503001 (50.42%)	2581 (0.26%) / 212614 (21.31%)	4950 (0.50%) / 194815 (19.53%)	7207 (0.72%) / 66826 (6.70%)	1084 (0.11%) / 2347 (0.24%)	 	good
		4,5×4,5	0.950	7518 (0.75%) / 722206 (72.39%)	3517 (0.35%) / 134010 (13.43%)	4668 (0.47%) / 77444 (7.76%)	2332 (0.23%) / 43476 (4.36%)	0 (0.00%) / 2467 (0.25%)	 	unsuccessful
		6×6	0.946	11511 (1.15%) / 729813 (73.15%)	2716 (0.27%) / 130353 (13.07%)	2233 (0.22%) / 79496 (7.97%)	1473 (0.15%) / 36215 (3.63%)	102 (0.01%) / 3726 (0.37%)	 	unsuccessful
		7,5×7,5	0.920	14168 (1.42%) / 845474 (84.75%)	3153 (0.32%) / 111917 (11.22%)	502 (0.05%) / 18674 (1.87%)	207 (0.02%) / 3203 (0.32%)	5 (0.001%) / 335 (0.034%)	 	unsuccessful
	8d – rich swamp forest	3×3	0.955	353 (0.04%) / 940490 (94.27%)	184 (0.02%) / 35933 (3.60%)	211 (0.02%) / 13893 (1.39%)	169 (0.02%) / 5966 (0.60%)	15 (0.002%) / 424 (0.04%)	 	unsuccessful

Rare vegetation types	8d – rich swamp forest	4,5×4,5	0.926	148 (0.01%) / 828532 (83.05%)	115 (0.01%) / 114081 (11.44%)	281 (0.03%) / 41809 (4.19%)	345 (0.03%) / 10742 (1.08%)	43 (0.004%) / 1542 (0.15%)			unsuccessful
		9×9	0.992	486 (0.05%) / 975335 (97.76%)	83 (0.01%) / 10500 (1.05%)	110 (0.01%) / 6305 (0.63%)	174 (0.02%) / 3778 (0.38%)	79 (0.01%) / 788 (0.08%)			unsuccessful
	9d – mud- bottom fens and bogs	3×3	0.989	3019 (0.30%) / 951032 (95.33%)	1373 (0.14%) / 22150 (2.22%)	1142 (0.11%) / 11570 (1.16%)	929 (0.09%) / 5337 (0.53%)	499 (0.05%) / 587 (0.06%)			unsuccessful
		4,5×4,5	0.993	4042 (0.41%) / 967591 (96.99%)	1038 (0.10%) / 12499 (1.25%)	765 (0.08%) / 6248 (0.63%)	666 (0.07%) / 3413 (0.34%)	451 (0.05%) / 925 (0.09%)			unsuccessful
		7,5×7,5	0.994	5014 (0.50%) / 960186 (96.25%)	878 (0.09%) / 15899 (1.59%)	445 (0.04%) / 9484 (0.95%)	327 (0.03%) / 4230 (0.42%)	298 (0.03%) / 877 (0.09%)			unsuccessful
		9×9	0.993	4042 (0.41%) / 967591 (96.99%)	1038 (0.10%) / 12499 (1.25%)	765 (0.08%) / 6248 (0.63%)	666 (0.07%) / 3413 (0.34%)	451 (0.05%) / 925 (0.09%)			unsuccessful

5.3 Predicted and true distribution on the map

The raster maps from MaxEnt's output, that show the predicted distribution of the vegetation types, allows a visual interpretation of the models precision that vary according to the grid mesh density of a Primary Statistical Unit (PSU). Figure 6 includes five maps of a small area located between Flaksjøen and Bølvatnet Lake. These maps show the visual comparison of predicted and true distribution of dwarf shrub heath (2e) from five PSU grid densities. Only the first two models have been interpreted as good models (Table 8). From the statistical analyses of MaxEnt results (Table 7), there was no correlation between the reduction of a grid mesh density and increasing probability values (RPPP) of the models. The highest RPPP value among these models belongs to the fourth model (0.90 in 0-1 scale) and the third model has the lowest value (only 0.69), however both of them were interpreted as unsuccessful. The second model has the best result from the comparison analyses (Table 8) and was nearly equally good as the first model, but the map is more clear and cells with highest RPPP value are more close to true distribution.

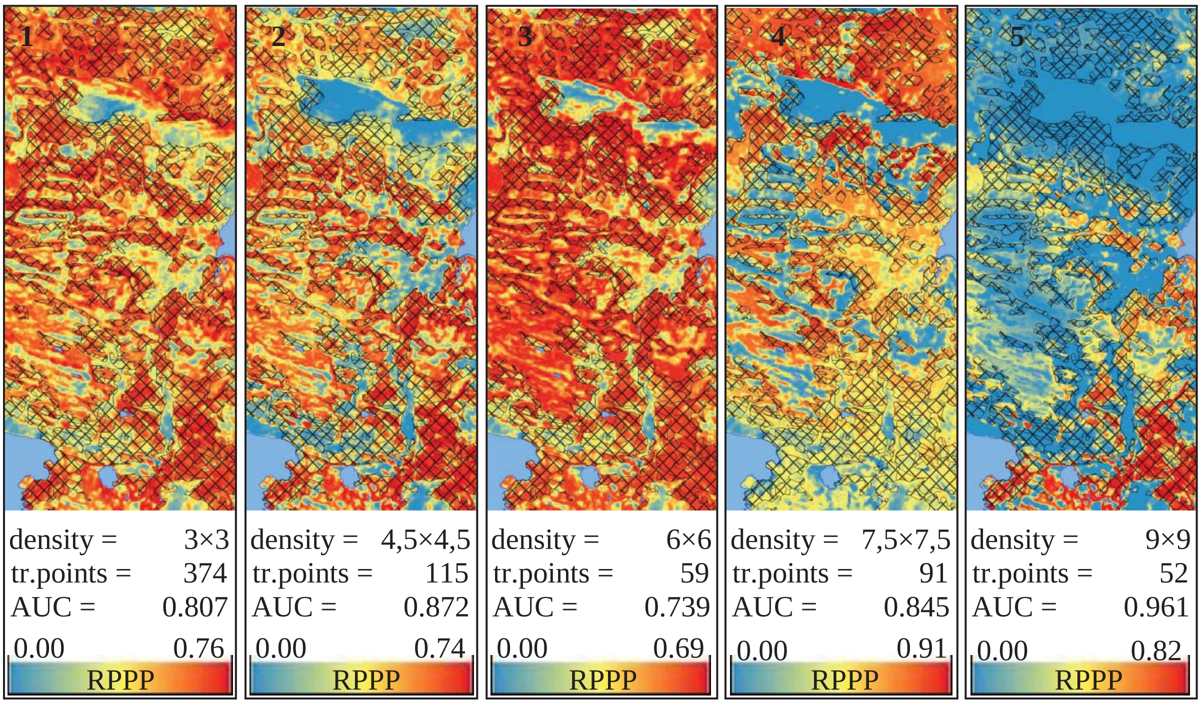


Figure 6. Visual representation of changing precision of MaxEnt models five maps. The figure shows the overlap of the predicted distribution (RPPP) in colors (from blue to red) and the true distribution in crosshatch of 2e vegetation type. Below the figure, the following information is provided: the density of PSU grid mesh used in the models, the summarized number of training points found in the sample units, AUC-values that show the MaxEnt models performance, as well as the highest RPPP value for each model.

The fifth model has the least precise result, and the true distribution is mostly covered by cells with low RPPP. The third model had only 7 training points less than the fifth model, but the result differs significantly. The reason for this difference is that training points (PO data) of the fifth model was taken from only one PSU unit, but the third model has got training points from three PSU units, located in geographically and ecologically different places. Therefore, it is important to note that the precision of the MaxEnt models depends not only on how many training points that are used for the modeling, but also the distribution of these points. The precision vary quite much among all five models in the study area, and are without doubt more accurate close to the training points. More detailed maps with comparison of predicted and true distribution are attached in Appendix 5.

6. Discussion

6.1 Implemented vegetation map

Based on the conducted fieldwork, the distribution of vegetation types on the implemented vegetation map (Bryn & Rekdal 2002) corresponds to the real distribution in 185 of 220 observation points (Table 6). This constitutes 84 % of all observation points. This gives the reasons to consider that the vegetation map from 2001used in the prediction modeling has a quality which if good enough for the purposes of the presented study.

The cause of insignificant errors in the implemented map could be the development of the vegetation types during the last 12 years or incorrect navigation caused by the difficulty of landscape, tree layer, weather and quality of the GPS receiver. The average GPS accuracy uncertainty during measurement of coordinates was 6.97 m, but varied from 5 m to 21 m. In addition, all observation points for the fieldwork were generated in ArcGIS with random spreading. Many of these points were generated close to the boundaries between vegetation types, and could thus be caused by difference in spatial resolution among the two very different approaches. The possible reason for significant errors on the implemented vegetation map is most probably human failure during vegetation mapping.

6.2 MaxEnt models

The main conclusion from the prediction modeling is that common vegetation types cannot be well modeled using a low PSU grid mesh densities, and the tested densities are not suitable for modelling of rare vegetation types. The main reason for this is probably the shortage of presence-only data (PO data) that was provided to present the modeled ecological and geographical variation of the vegetation types. In a study from Valdres, South-Central Norway, Ullerud (2013), found that some of the same vegetation types could be well modeled using MaxEnt with many of the same environmental variables. However, Ullerud (2013), used all presence locations for a test of model transferability (Randin et al. 2006), and therefore had many more PO points to train the MaxEnt models with (Hernandez et al., 2006). Also, as noted in the results, the distribution of the PO points is important for the model performance. Jimenez-Valverde et al. (2013) found that the distribution of the PO points was important for assessing the model performance, and that good models can only be achieved when the PO data are representative for the environmental variation within the study area. In the presented results, the obvious interpretation of low model performance for models run with a low density of PSU grid, should thus be that the PO points given by these PSU grids, are not representative for the environmental variation within the study area. The other reason that could affect the model performance can be the lack of more specified environmental variables and errors within the actual vegetation map. Most of the available environment variables used in the MaxEnt modeling describes primarily the abiotic environment. But it is recomended to include biotic interactions for species modeling, and this could potentionally also influence modeling of vegetation types. For example, in a study from Finland, Heikkinen et al. (2007) used the distribution of woodpecker species to predict owl distributions since woodpeckers provide nesting sites for owls by making cavities in trees.

Based on the MaxEnt evaluation of prediction models, the accuracy of models is greater for the rare vegetation types with more restricted environmental range and shown with high AUC values. This result has also been demonstrated by other comparable studies (Chahouki et al. 2010; Hernandez et al. 2006; Phillips et al. 2006). The common vegetation types had lower AUC values as than usually found (Phillips & Dudík 2008). Despite that almost all MaxEnt models was performing well according to the AUC values, the statistical analyses with evaluation data resulted in only 7 MaxEnt models interpreted as good models. It means that the MaxEnt models can be interpreted as successful at a first glance, but also that AUC values is not a good indicator of true model performance (Merow et al. 2013). Therefore, MaxEnt models with a high AUC value can in fact perform poorly when confronted with independent evaluation data (Halvorsen 2013).

Actually, this study supports that MaxEnt can be considered as a useful method for modeling the distribution of vegetation types, as found by Hemsing & Bryn (2012) and Ullerud (2013), but that it is of vital importance to confront the modelling results with independent evaluation data to assess the true performance, and furthermore that it is important to have PO training points that cover the entire range of environmental variation within the study area (Jimenez-Valverde et al. 2013).

The selection of environmental variables is crucial for the prediction modeling and often needs expert knowledge (Guisan & Zimmermann 2000; Manel et al. 2001). The MaxEnt ability to test variable importance using the "Jack-Knife test" allows to select the most important predictor variables and in turn to improve the model performance, following Halvorsen (2012). Based on the results in Table 7, the environmental predictors contribute differently in each of the modeled vegetation types. This is not very strange, since the vegetation types represent different parts of the ecological space (Bryn 2008; Ullerud 2013). The most widely distributed vegetation types, 2e and 4b, are seemingly regulated by the same set of environmental predictor variables. Other vegetation types that are less abundant within the study area, are regulated by other sets of predictors, and slope is becoming the most common environmental variable for the given distribution. Elevation is the most important environmental predictor variable that regulates the spatial distribution of all modeled vegetation types. This is clearly seen from the distribution of the vegetation types 2e and 4b, where 2e cover mostly mountainous areas whereas 4b is located mostly in lower parts. The use of satellite images in the modeling is justified by the fact that the different bands contribute to much of the model performance. The use of satellite images thus implies a great potential for further development within distribution modelling of vegetation types. From the satellite images, the blue band was the most important predictor variable for almost all modeled vegetation types (except 9c). The development of technical tools provided the opportunity to generate derived environmental variables, such as the NDVI (derivative of satellite image), TWI and slope (derivative of DEM), where NDVI was the most important predictor variable for MaxEnt modeling of forest and mountainous vegetation communities (2e, 3b, 4b and 8d). Increasing number of derived variables itself enriches the basis for DM and helps to investigate the influence of different factors of the distribution of vegetation types in a large variety of environmental parameters.

In this study, it was used a Primary Statistical Unit (PSU) as the sample unit for collection of the PO data. During preparation of the PO data, it was demonstrated that both shape and size of this unit are well suited to obtain enough PO data. The change of design of these units will probably influence the predictive performance of the models and provide changes in the modeling results (Stokland et al. 2011), because it would increase or decrease the number of training PO points for many of the vegetation types, and also change the distribution of the environmental predictor variables. Therefore it is probably interesting to look into the effect of prediction performance using different form of sample units (such us circles, routes, crosses, line grid) and changing their sizes. Another alternative for testing of models in different PSU grid densities can be the choice of representative sample units based on topographic features, because the geographical representation of sample units is probably more important than the number of samples, following Hengl et al. (2009).

The prediction modeling has shown that the use of different density in a point grid of presences for common (100×100m) and rare (25×25 m) vegetation types probably was a useful approach. However, these densities are probably not high enough. Table 7 shows that with increasing PSU grid density the number of training point increases significantly. The rare vegetation types had no PO data for some greater PSU grid densities. Close to the lowest PSU grid density, many modeled vegetation types had less than 100 training points collected in the whole study area. Thus, it makes sense to use higher density in a point grid of presences, especially for rare vegetation types that are limited by more specific environmental conditions. On the other hand, the inclusion of more PSU units is probably more important, since more PSU units will provide a better distribution of the PO points. Also, including more PO points for the common vegetation types, will increase the repeatability of the environmental variation, and thus only supply the modelling with redundant PO points, and thus slow down the running time for modelling in MaxEnt.

6.3 Geostatistical analysis

The geostatistical analyses that were used in the presented study are probably valid. But, during preparation of the data, some technical challenges arose in the collection of statistical data. The data was obtained as a point grid converted from a raster map that should be compared with true distribution in the form of a vector map. Points were located in the center of each raster cell and gave one value for a plot of 10×10 m size. During the joining of vegetation data onto this point grid, many points got zero values instead of codes for vegetation types. The reason for this is that these points were inserted in areas between two vegetation type polygons. Therefore, the vegetation map had to be converted to raster format in the same cell size and position as the raster map with the predicted distribution. In this way, the boundaries between the polygons have been deleted, but at the same time the spatial precision was reduced (Figure 7). However, the modeling was carried out on fairly large datasets, so it would probably not influence the results very much. Areas with high levels of human disturbance (cultivated lands and pastures) were excluded from this analysis, because MaxEnt will not perform very well for land cover types that are only indirectly explained by the available environmental variables (Hemsing 2010).

A possible alternative for this type of analysis, could be to overlay two maps with predicted and true distribution, previously converted to common format (vector or raster), and then to calculate the overlapped area and the rest area of the predicted and true distribution. The differences between these areas can be used further in evaluating and comparison of models. One of the preconditions here is that a map that shows predicted distribution will include polygons with only high RPPP values.

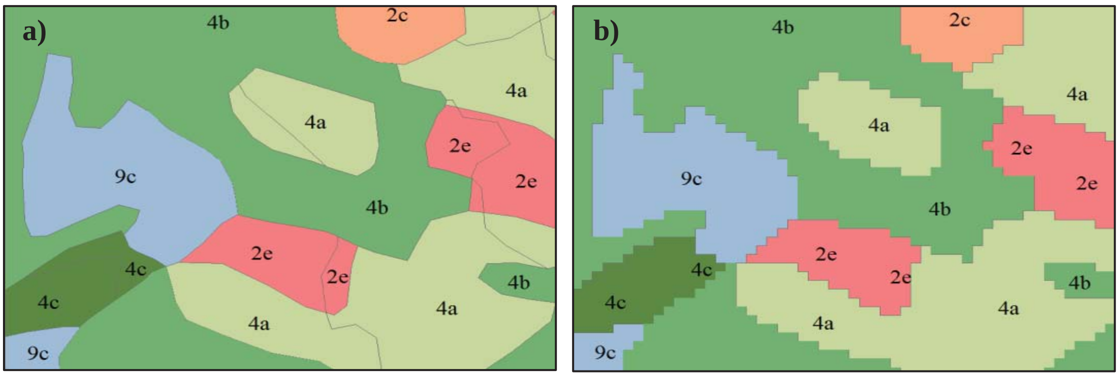


Figure 7. Visual presentation of changes in form and size of polygons on a map during conversion of vegetation map from vector (a) to raster format (b).

The modeling was executed with default settings in MaxEnt. Changing of some parameters as regularization multiplier and number of background points did not lead to significant improvements in the models. And the use of special settings needs adjustment of each model separately that subsequently would provide unequal settings among the MaxEnt models. But it opens great opportunities for research regarding the effects of varying settings on the performance of prediction models.

The PO data of selected vegetation types that belong to mosaics and make up less than 50 % of the cover within them were excluded from the modeling. The reason for this decision was to get the most correct ecological niches where modeled vegetation type can exist and were they do not overlapped or are mixed up with other vegetation types. On the other side, these mosaics show actual presences and can therefore be used as valuable material in the MaxEnt models as well as in the geostatistical comparison analyses. It was proven by several models that the predicted distribution is located to areas covered by mosaic polygons where modeled vegetation type was not dominant. Figure 8 show that in many places that are covered by mosaics, where 2e is secondary vegetation type, the probability of presence is very high. This gives reason to say that involvement of these mosaics into the modeling and model evaluation is very important and can improve the precision of prediction models.

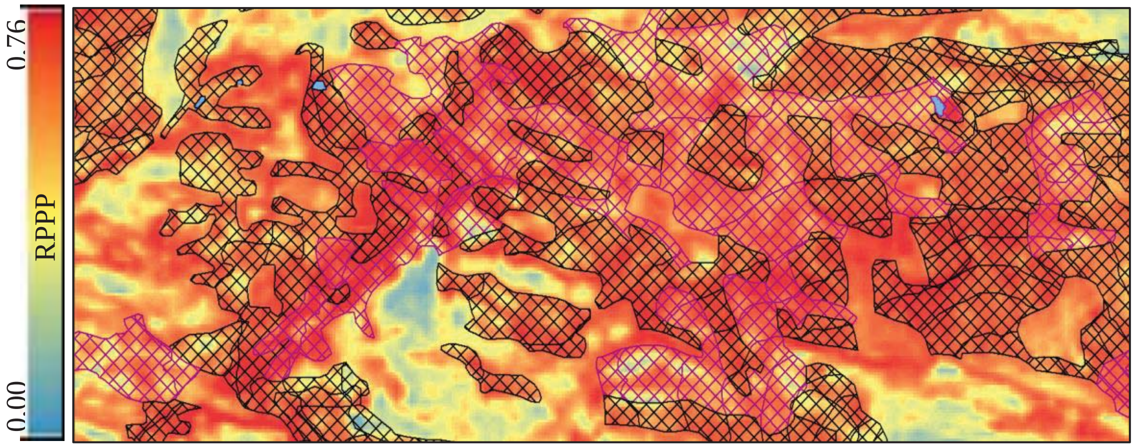


Figure 8. Comparison of predicted (background) and true (black crosshatch) distribution of 4b vegetation type included mosaics (purple crosshatch) where 2e cover less than 50 % of area (secondary vegetation type in a mosaic polygon).

The nature is a dynamic system that changes constantly as a result of various factors such as the species competition, climate changes, vegetation succession, invasive species, grazing, avalanches, wildfire, human influence and so on (Russell et al. 2011; Sala et al. 2000; Rosenzweig at al. 2007; Hernaux 1997; Bergeron and Archambault 1993, Weber and Flannigan 1997; Didier 2001). The interaction of these factors changes both the species composition within vegetation types and the species distribution area within the limits of environmental parameters. This can cause inaccuracies and challenges in the DM. The vegetation type is considered as a set of certain plant species that dominate both in tree, field and bottom layer (Rekdal & Larsson 2005). Many plant species have the same ecological niches and environmental conditions, while they belong to different vegetation types. Some plant species grow in several vegetation types. This can led to that the two or more vegetation types can have the spatial distribution limited by the same range of environmental parameters. As result of this “overlapping”, the MaxEnt can generate high RPPP in the areas outside the true distribution (Figure 8). Therefore it is reasonable first to model the distribution of the most important and dominant species separately and then by using overlapping to get summarized picture of modeled vegetation type, as done by Hemsing & Bryn (2012). On the other hand, we can consider these imprecisions on the map as potentially suitable distribution area for modeled vegetation types. This consideration can be used in the studies related to the potential natural vegetation types in areas strongly influenced by human activities, as in a study from Valders, Hemsing (2010).

7. Conclusion

The presented study has shown that distribution modeling of some vegetation types is possible using a frame area survey approach. It was obtained successful results for some vegetation types, indicating the clear relationships between modeled vegetation types and their environmental conditions, especially among common vegetation types in small densities of the PSU grid.

The collection of PO data was carried out in five densities of PSU grids for each vegetation type. Totally, of 30 planned models only 26 got PO data. Four models got no PO data and did therefore not contribute further in the prediction modeling. The PO data was obtained from the implemented vegetation map, which was structurally assessed for quality and errors before it was taken it use. The assessment was based on randomized field-observations within five areas of the map. The results showed that 84 % of the classified map corresponded with the real distribution.

Most of the prediction models were well evaluated by MaxEnt in according to the relative predictive ability. Of 26 tested models, 3 models have AUC-value less than 0.80 (more poor models). There was only one worthless model with an AUC-value of 0.500 (random model). It was not found any relationship between the number of presences in PO data and the AUC-value.

The comparison of the predicted and real distribution of modeled vegetation types has shown that only 6 of 26 prediction models can be considered as good models. Vegetation types 2e (dwarf shrub heath) and 4b (bilberry birch forest) were good modeled in 3×3 km and 4.5×4.5 km PSU grid densities. Vegetation type 3b (tall forb meadow) was good modeled in only 3×3 km PSU grid density. The vegetation type 9c (fen) has equal results for 3×3 km and 7.5×7.5 km PSU grid densities. The PO data for 9c were randomly located in the same PSU of these two densities. The best modeled vegetation type is 4b in 3x3 km PSU grid density. The vegetation types 8d (rich swamp forest) and 9d (mud- bottom fens and bogs) were not modeled successfully in any PSU grid densities, although they had high AUC-values.

The most important environmental variable that contributed to the prediction ability of all the modeled vegetation types was the DEM (the digital elevation model), NDVI index (the Normalized Difference Vegetation Index), slope and satellite image in blue band. The analysis of the variable importance conducted by MaxEnt (Jack-knife test) has shown that these variables have the most useful information. But the relative importance varies between the vegetation types and different PSU grid densities. These results show that the data generated from LIDAR-data and the satellite images contribute greatly to the performance of the prediction modeling.

This study could be repeated using the different form and size of the PSU, less PSU grid densities, different densities of point grid of the PO data, varying starting coordinates of the grids and alternative comparison analyses.

Literature

Almås R., Gjerdåker B., Lunden K., Myhre B. & Øye I., 2004. Norwegian Agricultural History, Tapir Academic Press, Trondheim.

Austin M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling, 157: 101/118. CSIRO Sustainable Ecosystems, Australia.

Austin M.P., 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological modelling, 200: 1–19. CSIRO Sustainable Ecosystems, Australia.

Bakkestuen V., Erikstad L. & Halvorsen R., 2008. Step-less models for regional environmental variation in Norway. Journal of Biogeography, 35 (10): 1906-1922.

Bekkby T., Erikstad L., Bakkestuen V. & Bjørge A., 2002. A landscape ecological approach to coastal zone applications. Sarsia, 87 (5): 396-408.

Bergeron Y. & Archambault S., 1993. Decreasing frequency of forest fires in the southern boreal zone of Que´bec and its relation to global warming since the end of the "Little Ice Age". Holocene, 3: 255–259.

Bergersen O.F., 1993. Trekk av Ringebus naturhistorie. Hemgrenda, 11-26.

Braun-Blanquet J., 1965. Plant Sociology: The Study of Plant Communities. Hafner, London.

Bryn A., 2006. Vegetation mapping in Norway and a scenario for vegetation changes in a mountain district. Geographia Polonica, 79 (1): 41-64.

Bryn A., 2008. Recent forest limit changes in south-east Norway: Effects of climate change or re-growth after abandoned utilisation? Norwegian Journal of Geography, 62 (4): 251-270.

Bryn A., Dourojeanni P., Hemsing L.O. & O'Donnell S., 2013. A high-resolution GIS null model of potential forest expansion following land use changes in Norway. Scandinavian Journal of Forest Research, 28 (1): 81-98.

Bryn A. & Murvold B., 2003. Vegetasjon og beite på Venabygdsfjellet. Hemgrenda, Ringebu. Historielag, 27: 115-123.

Bryn A. & Rekdal Y. 2002. Vegetasjon og beite på Venabygdsfjellet. Rapport fra vegetasjonskartlegging. NIJOS-rapport, 8/02: 68 s. Norsk institutt for jord- og skogkartlegging, Ås. ISBN 82-7464-298-8.

Chahouki M.A.Z., Azarnivand H., Jafari M. & Tavili A., 2010. Multivariate statistical methods as a tool for model-based prediction of vegetation types. Russian Journal of Ecology, 41 (1): 84-94.

Crawford P.H.C. & Hoagland B.W., 2009. Can herbarium records be used to map alien species invasion and native species expansion over the past 100 years? Journal of Biogeography, 36: 651–661.

Didier L., 2001. Invasion patterns of European larch and Swiss stone pine in subalpine pastures in the French Alps. Forest Ecology and Management, 145 (12): 67-77.

Elith J., 2002. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In: Ferson S. & Burgman M.(eds). Quantitative methods for conservation biology. Springer, 3958 pp.

Elith J., Graham C.H., Anderson R. P., Dudík M., Ferrier S., Guisan A., Hijmans R.J., Huettmann F., Leathwick J.R., Lehmann A., Li J., Lohmann L.G., Loiselle B.A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J.McC., Peterson A.T., Phillips S.J., Richardson K., Scachetti-Pereira R., Schapire R.E., Soberón J., Williams S., Wisz M.S. & Zimmermann N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29 (2): 129-151.

Eurostat, 2003. The Lucas survey. European statisticians monitor territory. Office for Official Publications of the European Communities, Luxembourg.

Ferrier S., Watson G., Pearse J. & Drielsma M., 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. Species-level modelling. Biodiversity & Conservation, 11: 2275-2307.

Fielding A.H. & Bell J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/ absence models. Environmental Conservation, 24: 3849 pp.

Florinsky I.V., 2012. Digital terrain analysis in soil science and geology. Elsevier/Academic Press, Amsterdam, 379 pp. ISBN 978-0-12-385036-2.

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

Franklin J., 2009. Mapping species distributions: spatial inference and prediction, 320 pp. Cambridge University Press, Cambridge.

Fremstad E., 1997. Vegetasjonstyper i Norge. Temahefte, 12: 1-279. Norsk institutt for naturforskning, Trondheim. ISBN: 82-426-0784-2.

Graham C. H., Ferrier S., Huettmann F., Moritz C. & Peterson A.T., 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol Evol, 19: 497-503.

Graham, C. H., Moritz C. & Williams S.E., 2006. Habitat history improves prediction of biodiversity in a rainforest fauna. Proceedings of the National Academy of Sciences of the United States of America, 103: 632-636.

Gudjonsson G., 2010. Large-scale vegetation mapping in Iceland. In Bryn A., Dramstad W. & Fjellstad W. (eds). Mapping and monitoring of nordic vegetation and landscapes, Hveragerði, Iceland. Viten fra Skog og Landskap, 1/10: 52-54.

Guisan A. & Zimmermann N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling, 135 (2-3): 147-186.

Haines-Young R., 1991. Biogeography. Progress in Physical Geography, 15: 101-13.

Halvorsen R., 2012a. A gradient analytic perspective on distribution modelling. Sommerfeltia, 35: 1-165. Department of Research and Collections, Natural History Museum, University of Oslo, Olso. ISBN 82-7420-049-7.

Halvorsen R., 2012b. A maximum likelihood explanation of MaxEnt, and some implications for distribution modelling. Sommerfeltia, 36: 1-132.

Halvorsen R., 2013. A gradient analytic perspective on distribution modelling. Sommerfeltia, 35: 1-164.

Hemsing L.Ø. & Bryn A., 2012. Three methods for modelling potential natural vegetation (PNV) compared: A methodological case study from south-central Norway. Norwegian Journal of Geography, Volume 66: 11-29. ISSN: 0029-1951.

Heikinnen R.K., Luoto M., Virkkala R., Pearson R.G. & Körber J-H., 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. Global Ecology and Biogeography, 16: 754-763.

Hiernaux P., Fernandez-Revira S. & Williams T., 1997. Dynamics of livestock-mediated nutrient transfers in Semi-Arid West-Africa (SAWA) landscapes: implications on natural resources management. ILRI Annual Project Reports. International Livestock Research Institute, 86-88.

Hernandez P.A., Graham C.H., Master L.L. & Albert D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 29 (5): 773-785.

Hovdhauge E., 1988. Busetningshistoria for Venabygd. Hemgrenda, 12: 8–24.

Jiménez-Valverde A., Ecevedo P., Barbosa A.M., Lobo J.M. & Real R., 2013. Discrimination capacity in species distribution models depends on the representativeness of the environmental domain. Global Ecology and Biogeography, 22: 508–516.

Küchler A.W. & Zonneveld I.S. (eds), 1988. Vegetation Mapping. Kluwer, Dordrecht.

Larsson J.Y. & Rekdal Y., 1997. Veiledning i vegetasjonskartlegging M 1:50 000. Dokument, 5-97. Norsk institutt for jord- og skogkartlegging, Ås.

Lawesson J.E., Eilertsen O., Diekmann M., Reinikaien A., Gunnlaugsdottir E., Fosaa A.M., Carøe I., Skov F., Groom G., Økland T., Økland R., Anderson P-N. & Bakkestuen V., 2000. A concept for Vegetation Studies and Monitoring in the Nordic Countries. Tema Nord 517: 125 pp. Nordic Council of Ministers, Copenhagen. ISBN: 92-893-0434-0.

Manel S., Williams H.C. & Ormerod S.J., 2001. Evaluating presence-absence models in ecology: The need to account for prevalence. Journal of Applied Ecology, 38 (5): 921-931.

Martinez-Taberner A., Ruiz-Perez M., Mestre I. & Fortenza V., 1992. Prediction of potential submerged vegetation in a silted coastal marsh, Albufera of Majorca, Balcaric Islands. Journal of Environmental Management, 35: 1-12.

Merow C., Smith M.J. & Silander J.A., 2013: A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography, 36: 1-12.

Moen A., 1999. National Atlas of Norway: Vegetation. Norwegian Mapping Authority, Hønefoss, 200pp.

Moravec, J. 1998. Reconstructed natural versus potential natural vegetation in vegetation mapping: A discussion of concepts. Applied Vegetation Science, 1: 173-176.

Palmer A.R. & Van Staden J.M., 1992. Predicting the distribution of plant communities using annual rainfall and elevation: an example from southern Africa. Journal of Vegetation Science, 3: 261-66.

Parry R.B. & Perkins C., 2002. World mapping today. De Gruyter Saur. ISBN: 3-598-11534-2.

Pearce J. & Ferrier S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling, 133 (3): 225-245.

Phillips S.J., Anderson R.P. & Schapire R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190 (3–4): 231-259.

Phillips S.J. & Dudík M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography, 31: 161-175.

Phillips S.J, Dudík M. & Schapire R.E., 2004. A maximum entropy approach to species distribution modeling. In Proceedings of the Twenty-First International Conference on Machine Learning, 655-662.

Phillips S. J., Dudik M. & Schapire R.E., 2005. A brief tutorial on Maxent. AT&T Research: 38 pp.

Puschmann O., 2005. Nasjonalt referansesystem for landskap - Beskrivelse av Norges 45 landskapsregioner. NIJOS rapporter 10/05: 204 pp.

Randin C.F., Dirnböck T., Dullinger S., Zimmermann N.E., Zappa M. & Guisan A., 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography, 33: 1689-1703.

Rekdal Y. & Bryn A., 2010. Mapping of vegetation in Norway. Viten fra Skog og landskap 01/10: 93-96. Norwegian Forest and Landscape Institute, Ås.

Rekdal Y. & Larsson J.Y., 2005. Veiledning i vegetasjonskartlegging. M 1:20000–50000. NIJOS-instruks nr. 1/05. Norsk institutt for jord- og skogkartlegging, Ås. ISBN: 82-7464-341-0.

Remsen D., Ko B., Chavan V. & Raymond M., 2011. Getting Started: An overview of data publishing in the GBIF network. Copenhagen: Global Biodiversity Information Facility, 1-16. ISBN: 87-92020-28-3.

Rosenzweig C., Casassa G., Karoly D.J., Imeson A., Liu C., Menzel A., Rawlins S., Root T.L., Seguin B., Tryjanowski P. & Hanson C.E., 2007. Assessment of observed changes and responses in natural and managed systems. Cambridge University Press, 79-131.

Russell P., Hertz P. & McMillan B., 2011. The effects of population interactions on community characteristics. Biology: The Dynamic Science, Volume 2: 560 pp. ISBN: 978-0-538-49373-4.

Sala O.E., Chapin F.S., Armesto J.J., Berlow R., Bloomfield J., Dirzo R., Huber-Sanwald E., Huenneke L.F., Jackson R.B., Kinzig A., Leemans R., Lodge D., Mooney H.A., Oesterheld M., Poff N.L., Sykes M.T., Walker B.H., Walker M. & Wall D.H., 2000. Global biodiversity scenarios for the year 2100. Science, 287: 1770-1774.

Siedlicka A., Nystuen J.P., Englund J.O. & Hossack J., 1987. Lillehammer berggrunnskart, M 1:250 000. Norges geologiske undersøkelse.

Sillero N., Brito J.C., Martín-Alfageme S., García-Meléndez E., Toxopeus A.G. & Skidmore A., 2012. The significance of using satellite imagery data only in Ecological Niche Modelling of Iberian herps. Acta Herpetologica 7 (2): 221-237.

Sollid J.L. & Trollvik J.A., 1991. Oppland Fylke, kvartærgeologi og geomorfologi 1:250 000. Institutt for Naturgeografi, Universitetet i Oslo.

Stokland J.N., Bakkestuen V., Bekkby T., Rinde E., Skarpaas O., Thygeson A.S., Yoccoz N.G. & Halvorsen R., 2008. Prediksjonsmodeller som verktøy for kartlegging, overvåking og forvaltning av biologisk mangfold – anvendelse, utviklingspotensial og utfordringer. Rapport, 1: 1-72. Naturhistorisk museum, Oslo.

Stokland J. N., Halvorsen R. & Støa B., 2011. Species distribution modelling- Effect of design and sample size of pseudo-absence observations. Ecological Modelling, 222 (11): 1800-1809.

Strand G.H., 2013. The Norwegian area frame survey of land cover and outfield land resources. Norwegian Journal of Geography, Volume 67: 24-35. ISSN 0029-1951.

Strand G.-H. & Rekdal Y., 2010. The Norwegian area frame survey of land resources: AR18×18. Viten fra Skog og landskap, 01/10: 113-116.

Taugbøl T., 2002. Hyttebygging i Norge - Trender, konflikter og styringsmuligheter. Norsk institutt for naturforskning, www.utmark.no

Thomas C. D., Cameron A., Green R.E., Bakkenes M., Beaumont L.J., Collingham Y.C., Erasmus B.F.N., Siqueira M.F., Grainger A., Hannah L., Hughes L., Huntley B., van Jaarsveld A.S., Midgley G.F., Miles L., Ortega-Huerta M.A., Peterson A.T., Phillips O.L. & Williams S.E., 2004. Extinction risk from climate change. Nature, 427: 145-148.

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

Ullerud H.A., 2013. Vegetation type distribution modeling with independent data evaluation - a local-scale study from the boreal-alpine transition zone. Department of Ecology and Natural resource Management. Master Thesis. Norwegian University of Life Science, Ås.

Ward G., Hastie T., Barry S., Elith J. & Leathwick J.R., 2009. Presence-only data and the EM algorithm. Biometrics, 65: 554-563.

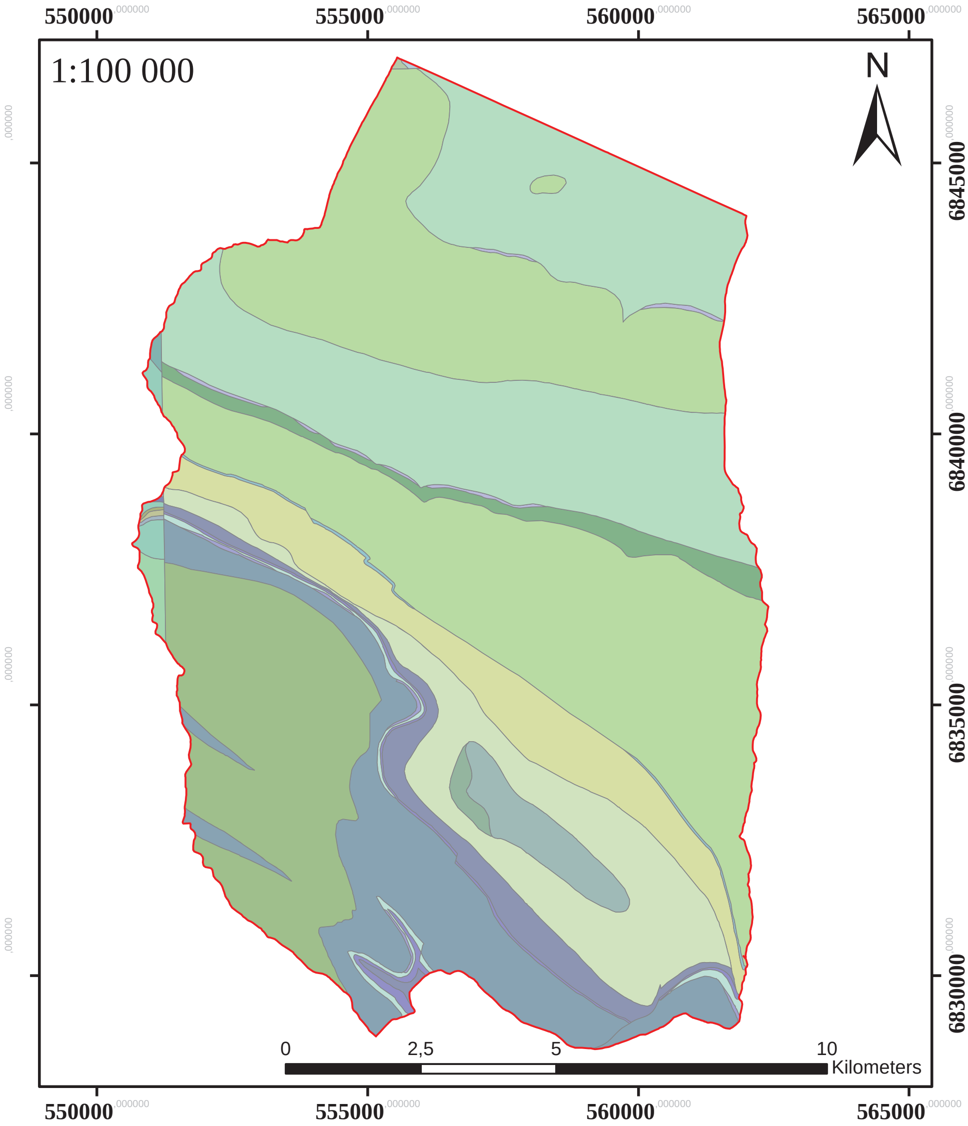
Weber M.G. & Flannigan M.D., 1997. Canadian boreal forest ecosystem structure and function in a changing climate: impact on fire regimes. Environmental Reviews, 5: 145–166.

Weier J. & Herring D., 2000. Measuring Vegetation (NDVI & EVI), http://earthobservatory.nasa.gov/Features/MeasuringVegetation/

Wollan A. K., Bakkestuen V., Kauserud H., Gulden G. & Halvorsen R., 2008. Modelling and predicting fungal distribution patterns using herbarium data. Journal of Biogeography, 35 (12): 2298-2310.

Appendices

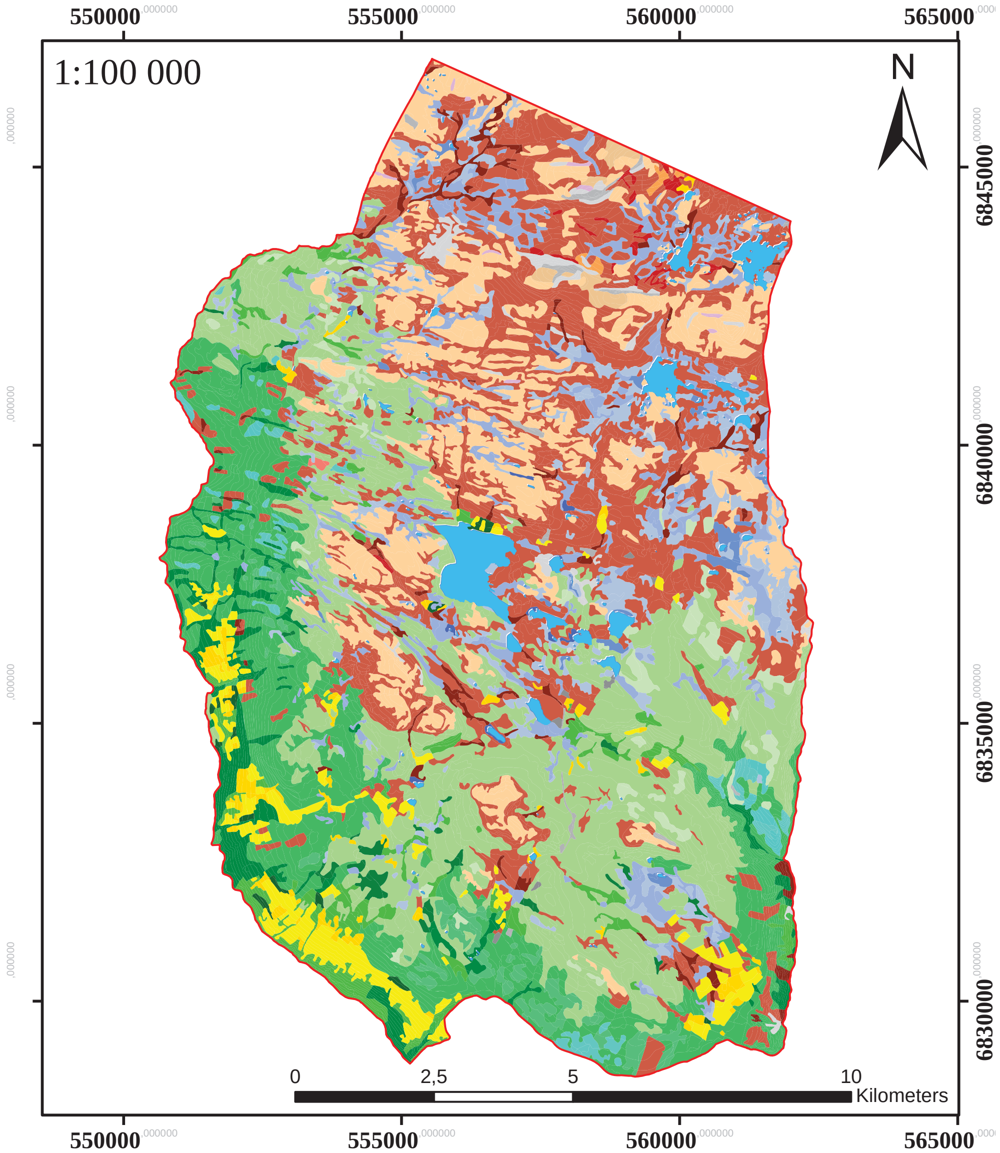
Appendix 1. Bedrock map of the study area.



Legend:

- Dolomite with layers of sandstone or sandy phyllite
- Dolomitic metasandstone
- Dolomic metasandstone and gray metasandstone with layers of phyllite
- Phyllite and phyllitic sandstone
- Schist, augen gneiss and metagabbro/green slate
- Gneiss and augen gneiss
- Granite, foliated
- Grey feldspathic metasandstones, in some places with feldspar
- Grey metasandstone
- Gray meta-sandstone with layers of conglomerate and phyllite
- Grey metasandstone with some layers of polymikt conglomerate, gray sandy phyllite
- Gray-green conglomerate with layers of sandstone and phyllite
- Light feldspathic sandstone
- Light gray feldspathic metasandstone and quartzite
- Light gray feldspathic metasandstone
- Light gray quartzite with varying content of feldspar
- Light gray metasandstone, sandy phyllite
- Metagabbro
- Metasandstone, unsorted
- Dark gray greywacke, black shale, conglomerate
- Dark gray, poor sorted sandstone (greywacke), black shale, conglomerate
- Sandstone
- Sandstone, conglomerate and phyllite
- Unsorted conglomerates
- Augen gneiss

Appendix 2. Vegetation map of the study area.



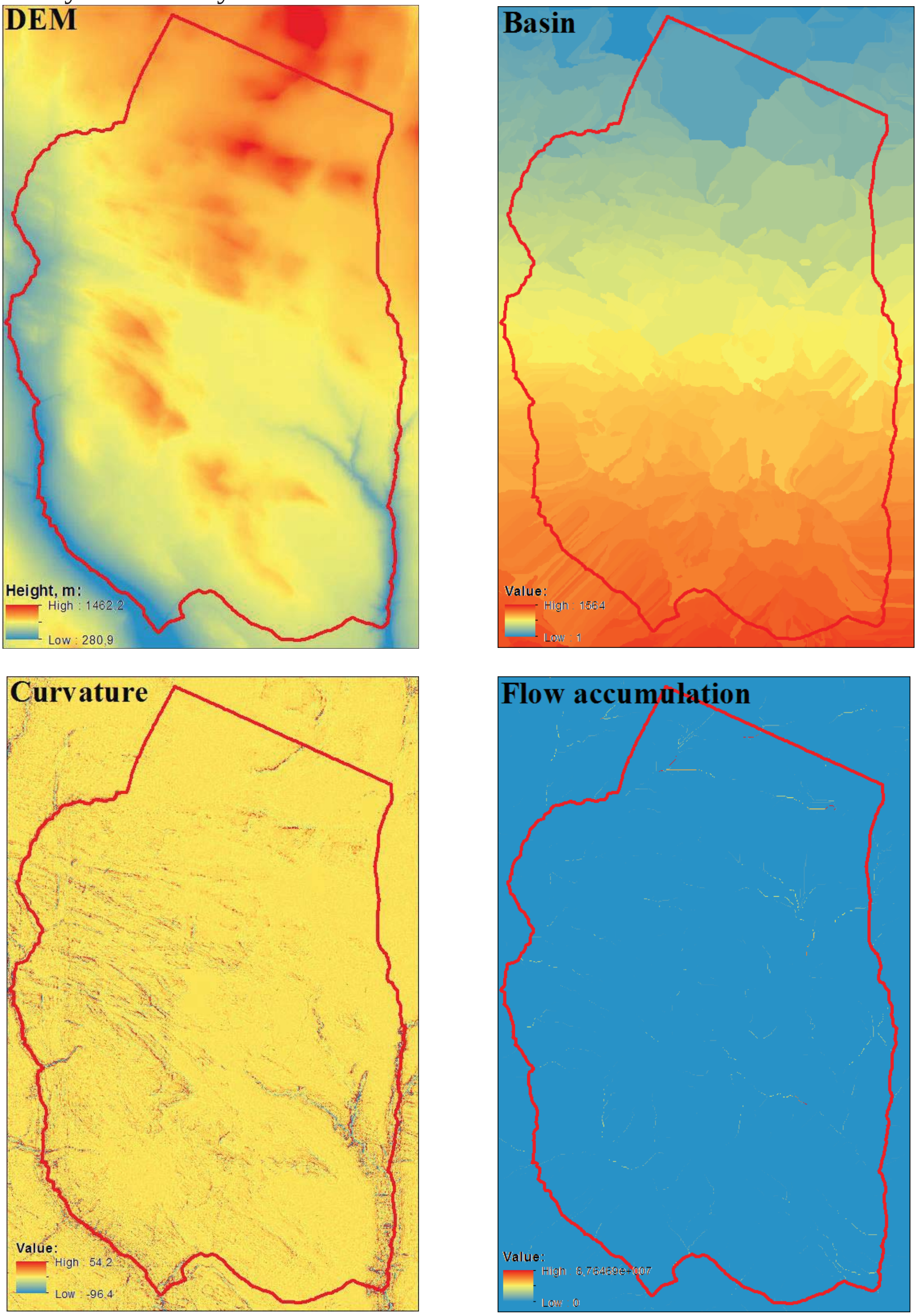
Legend:

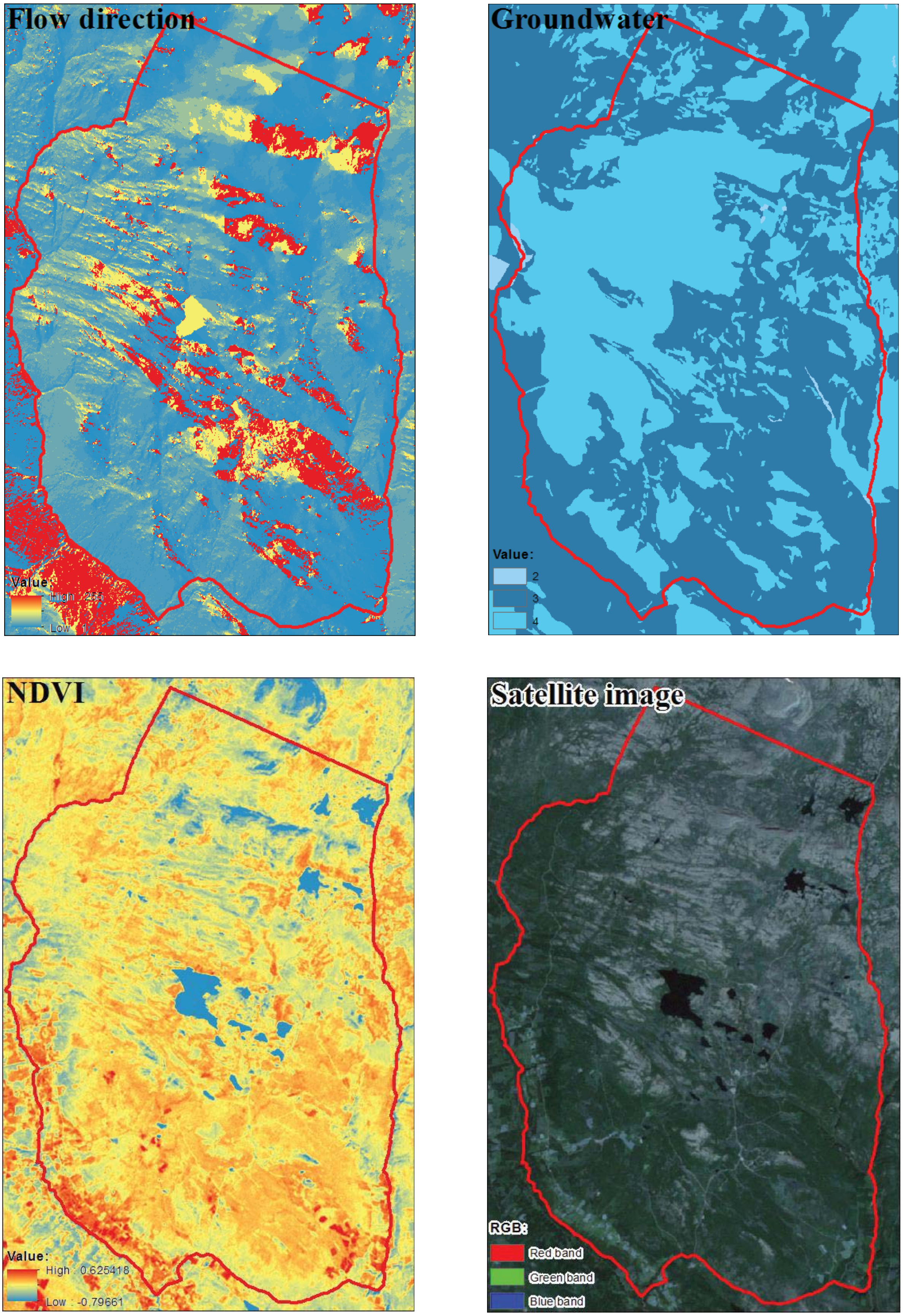
Moss snow-bed
Sedge and grass snow-bed
Stone polygon land
Mid-alpine heath
Dry grass heath
Lichen heath
Mountain avens heath
Dwarf shrub heath
Alpine calluna heath
Alpine damp heath
Low herb meadow
Tall herb meadow
Lichen- and heather birch forest
Blueberry birch forest
Meadow birch forest
Birch forest on lime soils
Alder forest
Flood-plain shrubs
Pasture land forest
Oak forest
Beech forest
Broad-leaved deciduous forest
Lichen- and heather pine forest
Blueberry pine forest
Meadow pine forest
Pine forest on lime soils
Lichen- and heather spruce forest
Blueberry spruce forest
Meadow spruce forest
Damp forest
Bog forest
Poor swamp forest
Rich swamp forest
Bog
Deer-grass fen
Fen
Mud-bottom fens and bogs
Sedge marsh
Coastal heath
Calluna heath
Damp heath
Moist meadows
Dunes and gravel beaches
Alluvial sand and gravel planes
Cultivated land
Pastures
Barren land
Boulder field
Exposed bedrock
Built-up areas
Scattered housing
Artificial impediment
Glacier
Not mapped
Water

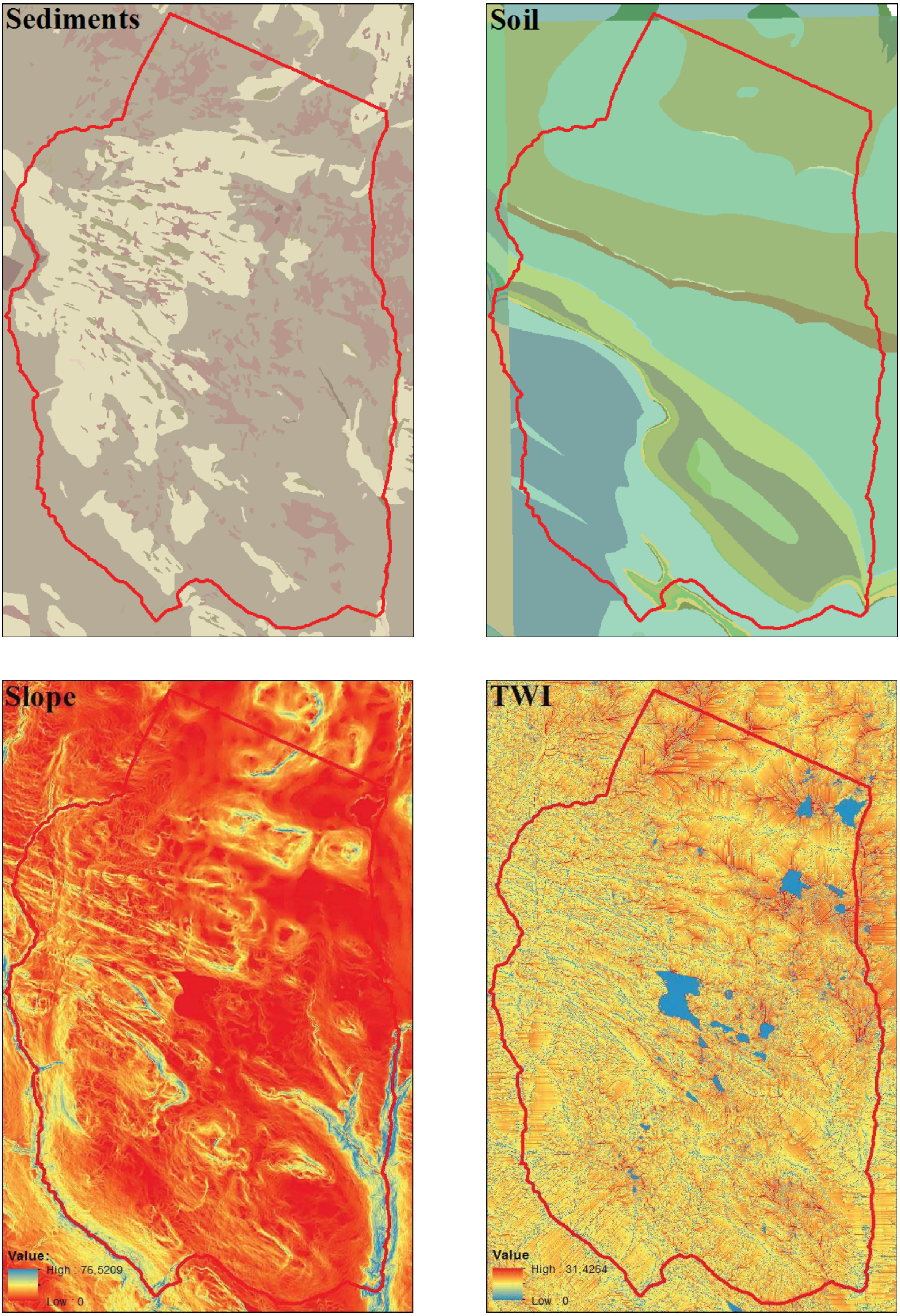
Appendix 3. Additional signs used to describe variation within the modeled vegetation types (table 3).

Additional signs	Described variation
o)	deciduous trees, unspecified
+	Scots pine
*	Norwegian spruce
j	more than 50 % cover of *Juniperus*
ɔ	25-50 % cover of willows
s	more than 50 % cover of willows
g	grass-rich vegetation
v	25 – 50 % cover of lichens
x	more than 50% cover of lichens
⟠	50-75 % cover of stone and block
]	25-50 % cover of trees
k	calcareous vegetation
n	more than 50 % cover of *Nardus stricta*
H	cut areas or young forest

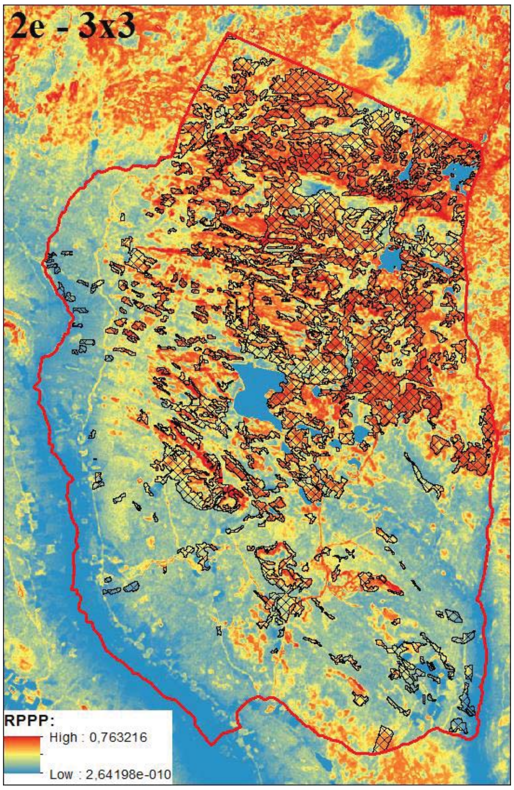
Appendix 4. Environmental variables. In the prediction modeling all variables were used as the raster maps. Below, the maps show the variety (in colors) of each environmental parameter in the study area. The red line shows the boundary of the study area.

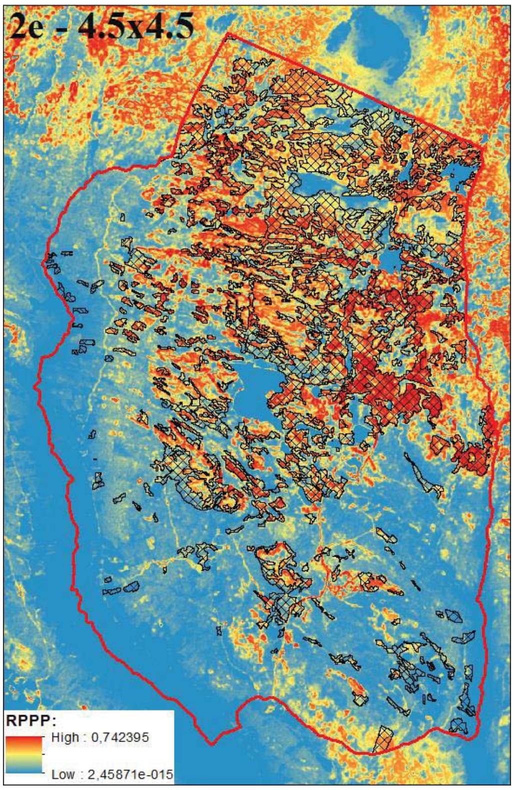


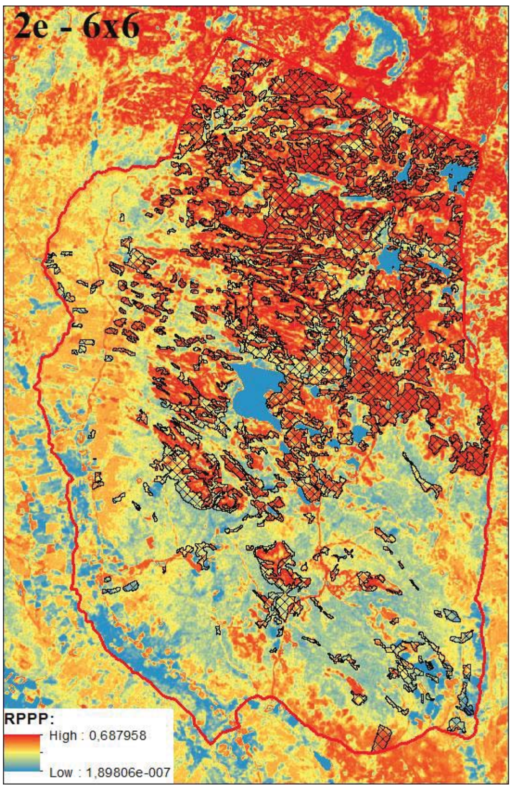

Flow direction
Value:
High : 255
Low : 1
Groundwater
Value:
2
3
4
NDVI
Value:
High : 0.625418
Low : -0.79661
Satellite image
RGB:
Red band
Green band
Blue band

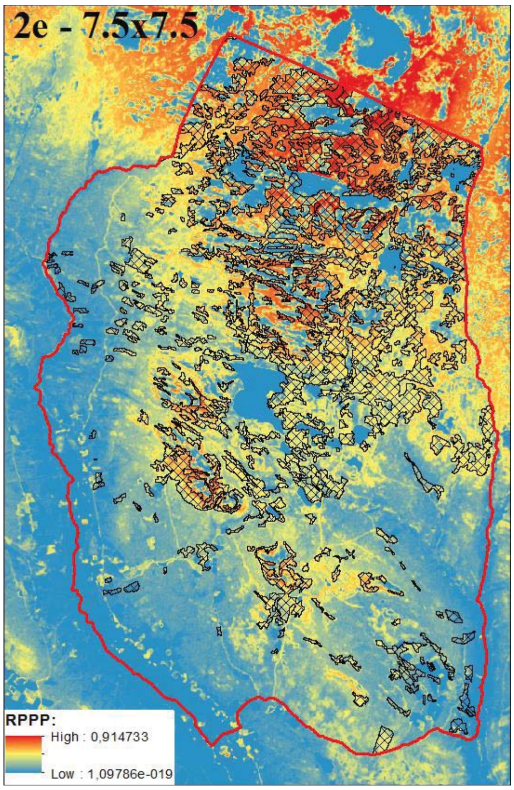

Sediments
Soil
Slope
Value:
High : 76.5209
Low : 0
TWI
Value
High : 31.4264
Low : 0

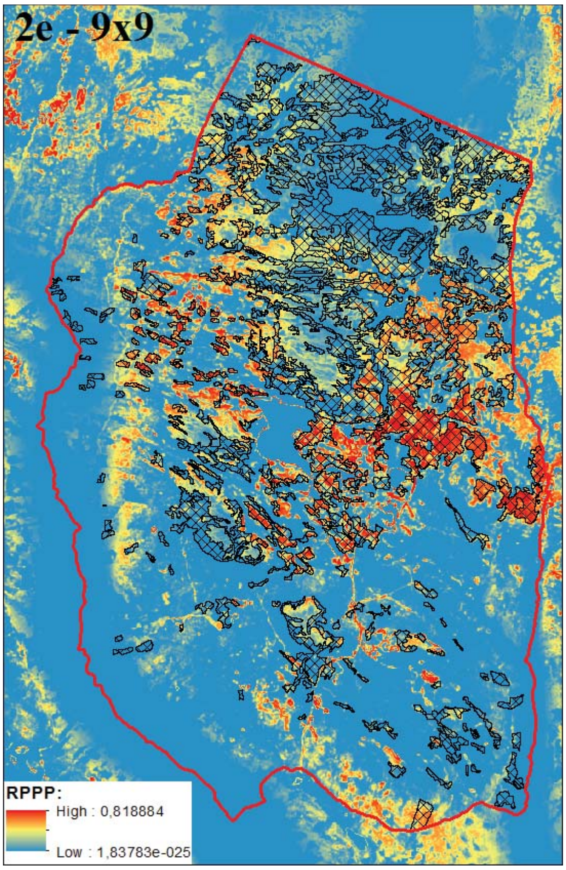
Appendix 5. Comparison analysis. The maps show the comparison of predicted and real distribution of modeled vegetation types. The predicted distribution is shown as a raster map in colors. Warmer colors show higher probability of presence. The red line shows the boundary of the study area. The black crosshatch shows the true distribution of the modeled vegetation types.

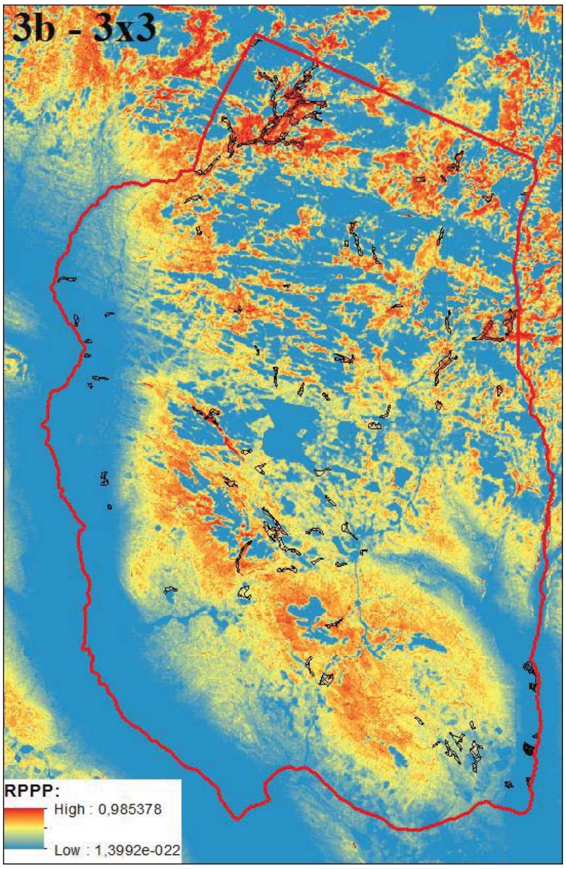


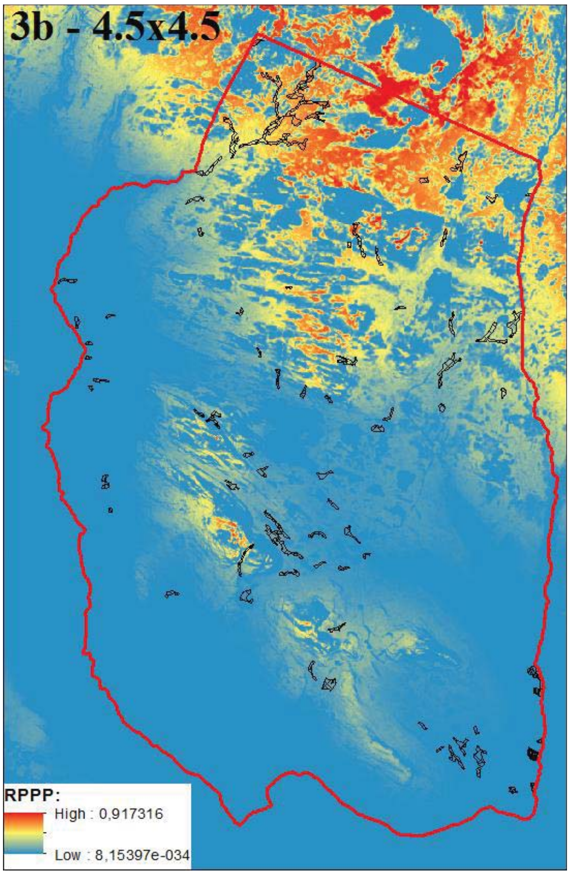


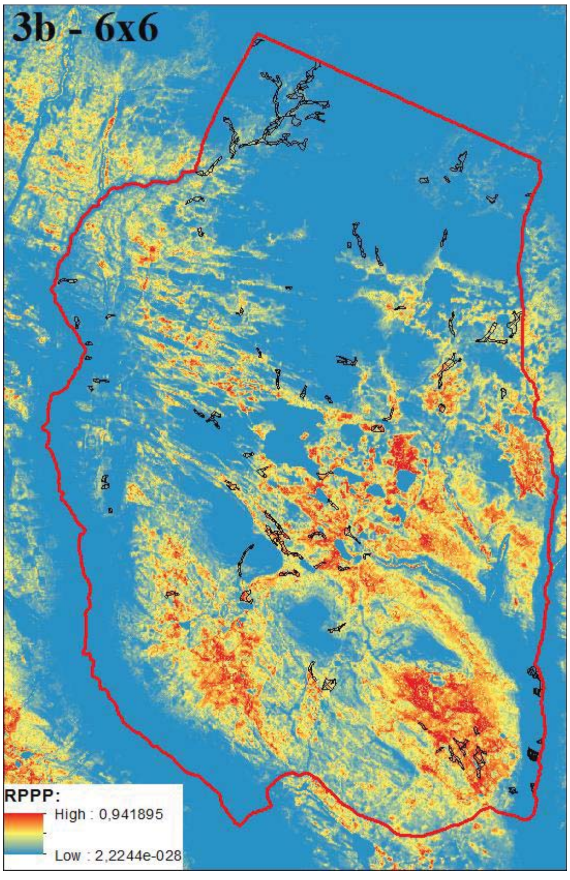


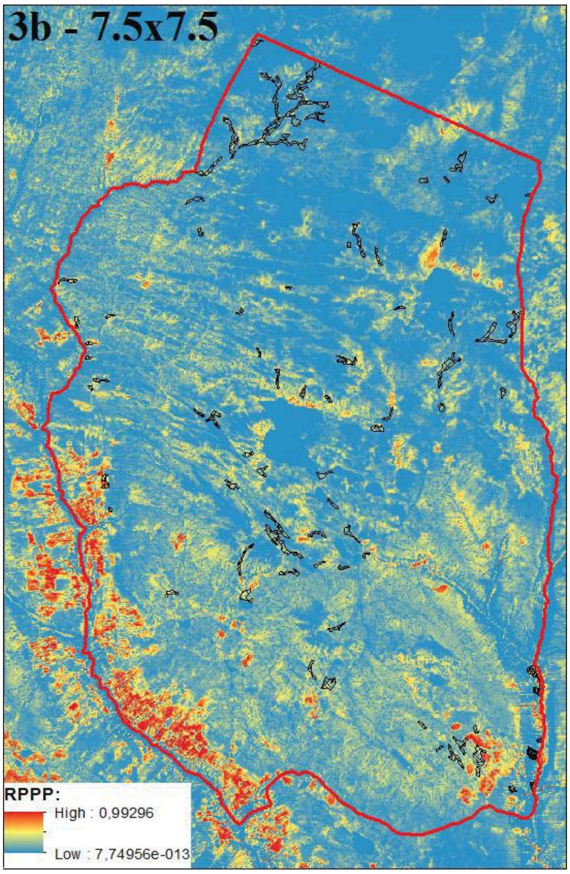


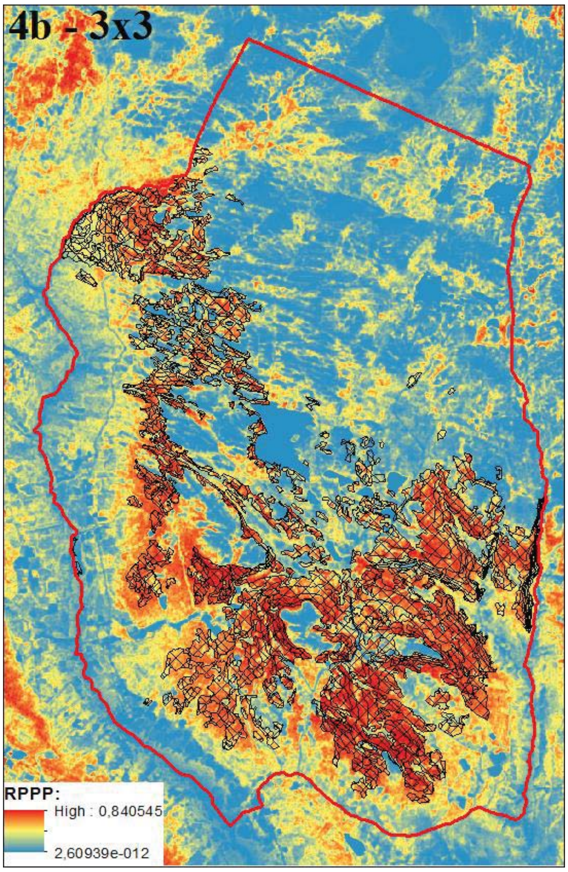

2e - 9x9
RPPP:
High : 0,818884
Low : 1,83783e-025

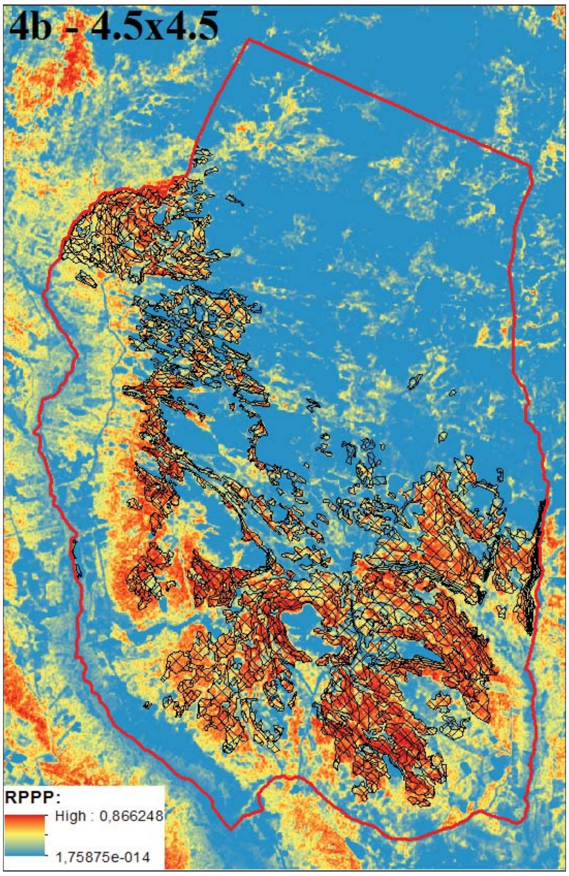

3b - 3x3
RPPP:
High : 0,985378
Low : 1,3992e-022

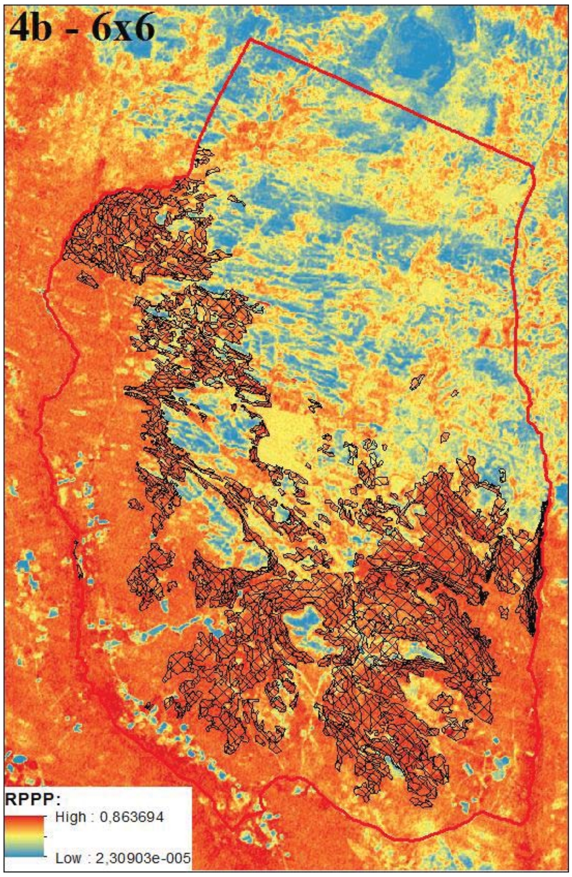

3b - 4.5x4.5
RPPP:
High : 0,917316
Low : 8,15397e-034

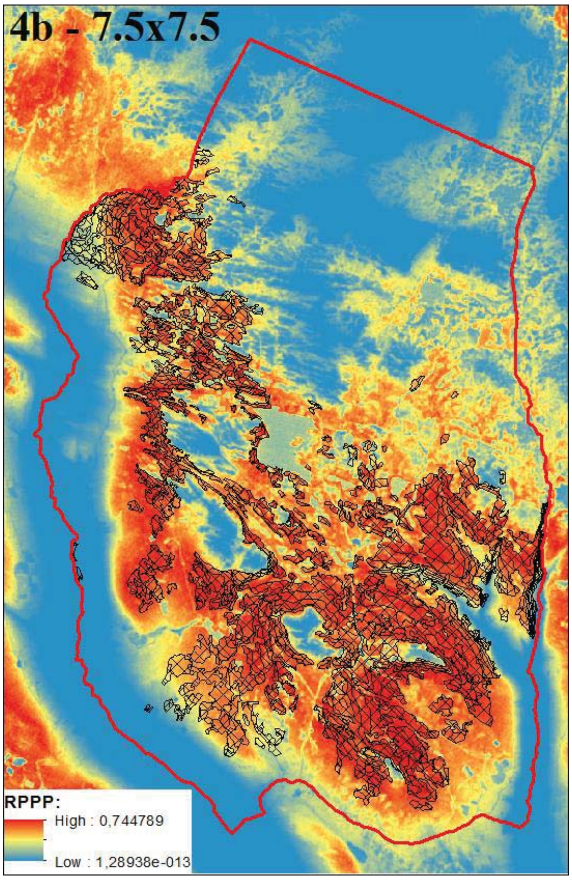

3b - 6x6
RPPP:
High : 0,941895
Low : 2,2244e-028

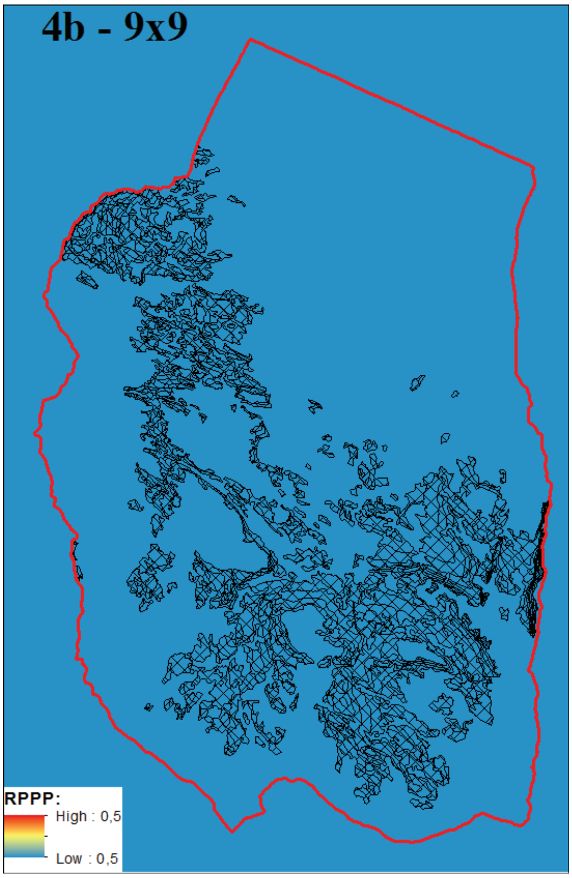

3b - 7.5x7.5
RPPP:
High : 0,99296
Low : 7,74956e-013

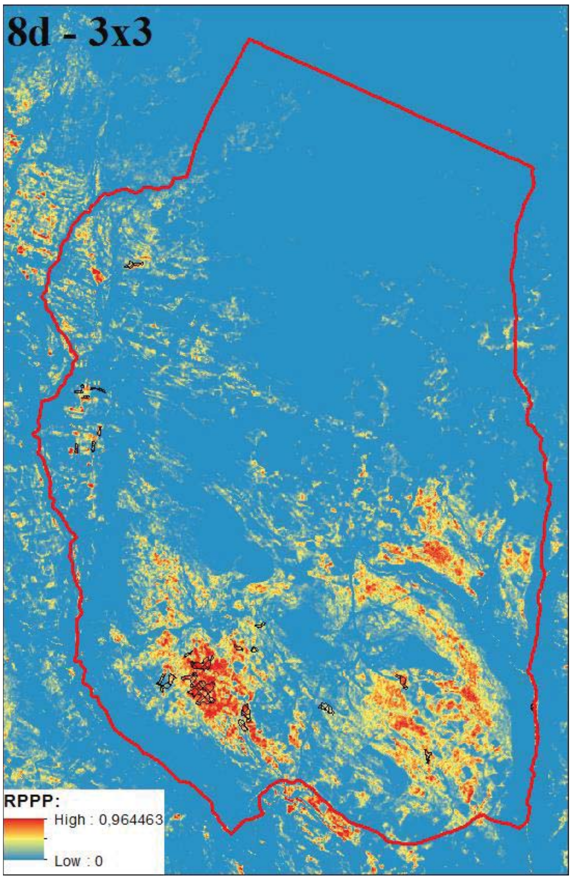

4b - 3x3
RPPP:
High : 0,840545
2,60939e-012

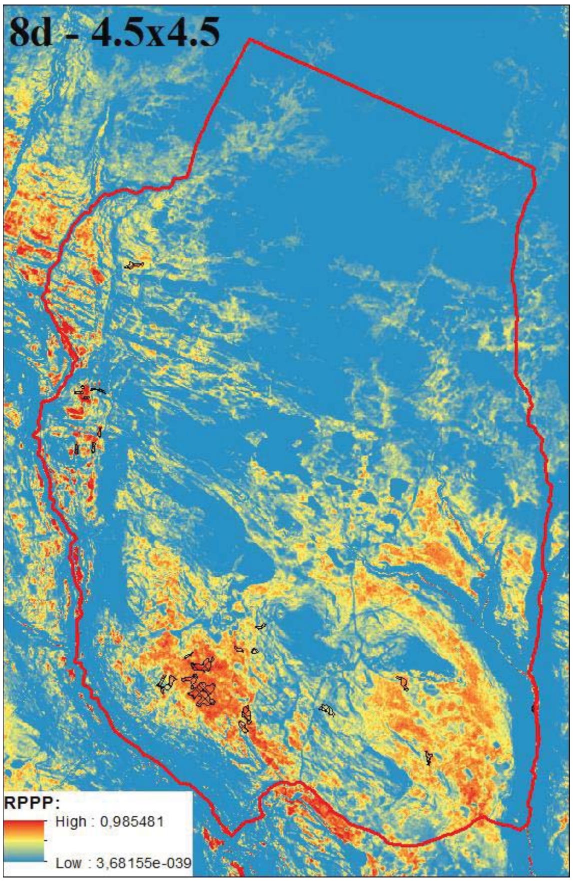

4b - 4.5x4.5
RPPP:
High : 0,866248
1,75875e-014

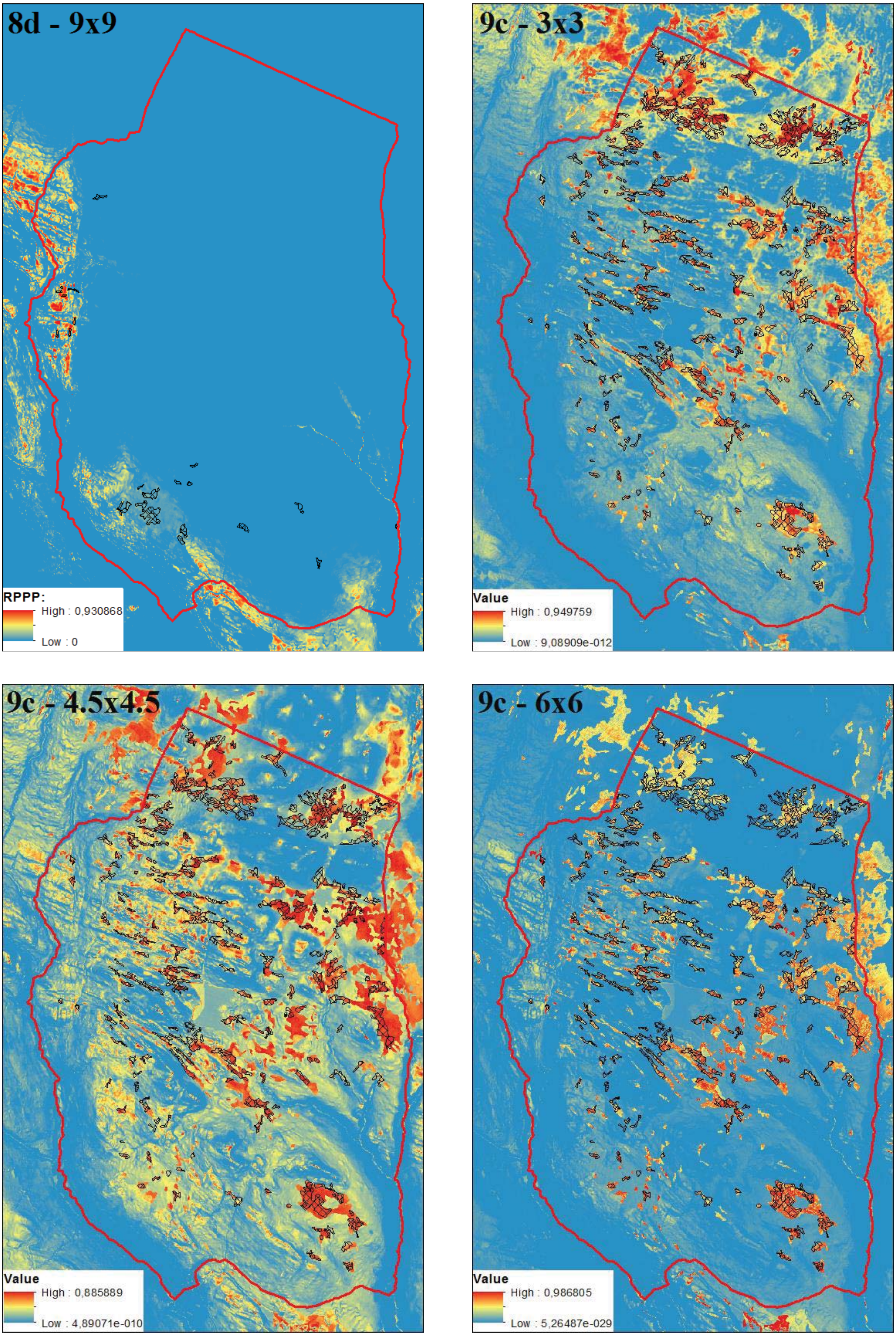

4b - 6x6
RPPP:
High : 0,863694
Low : 2,30903e-005

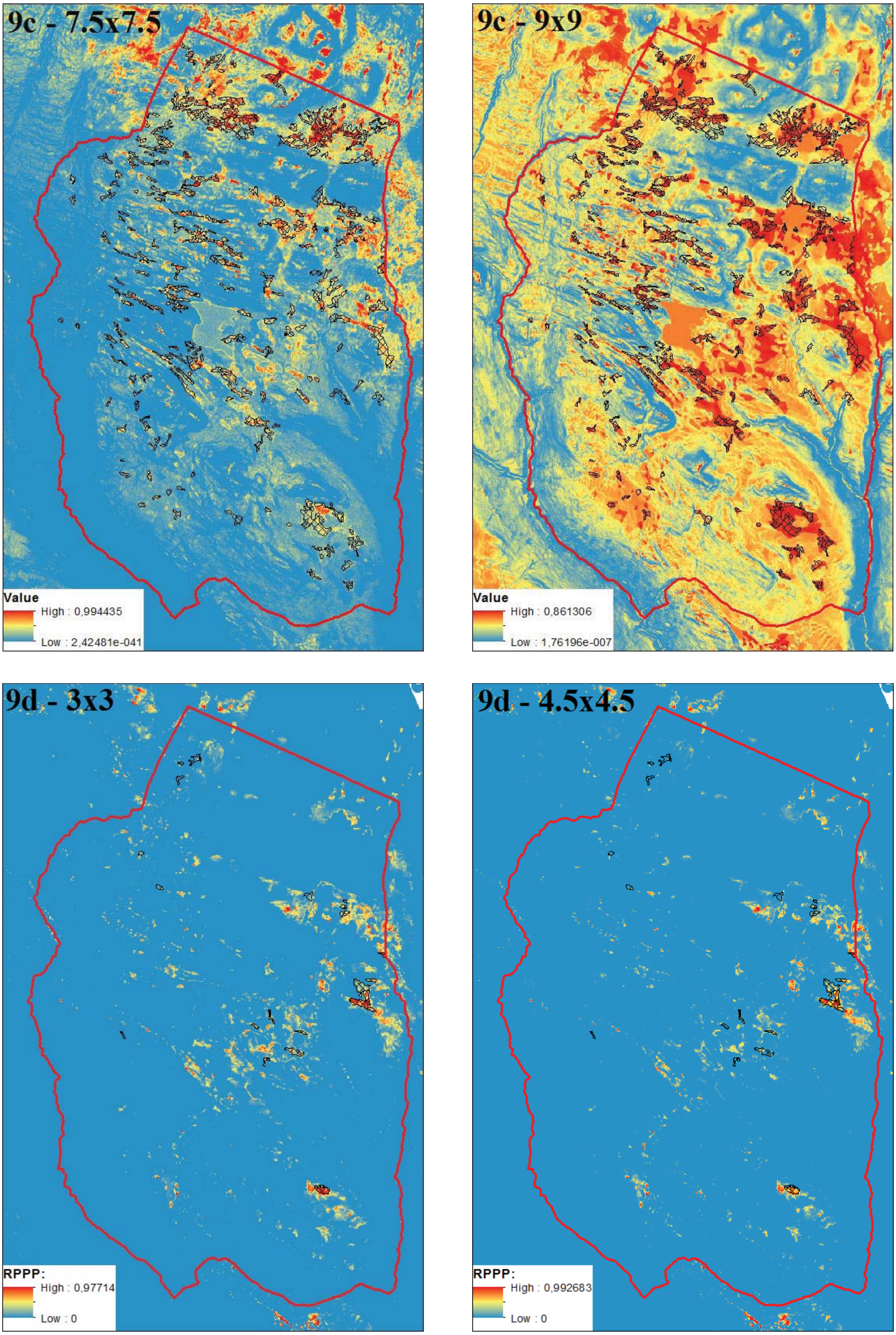

4b - 7.5x7.5
RPPP:
High : 0,744789
Low : 1,28938e-013

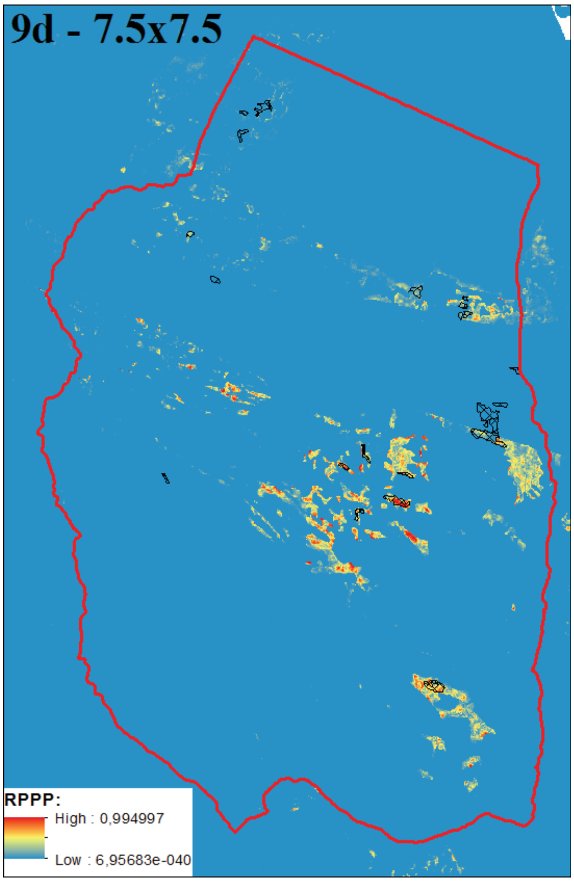

4b - 9x9
RPPP:
High : 0,5
Low : 0,5

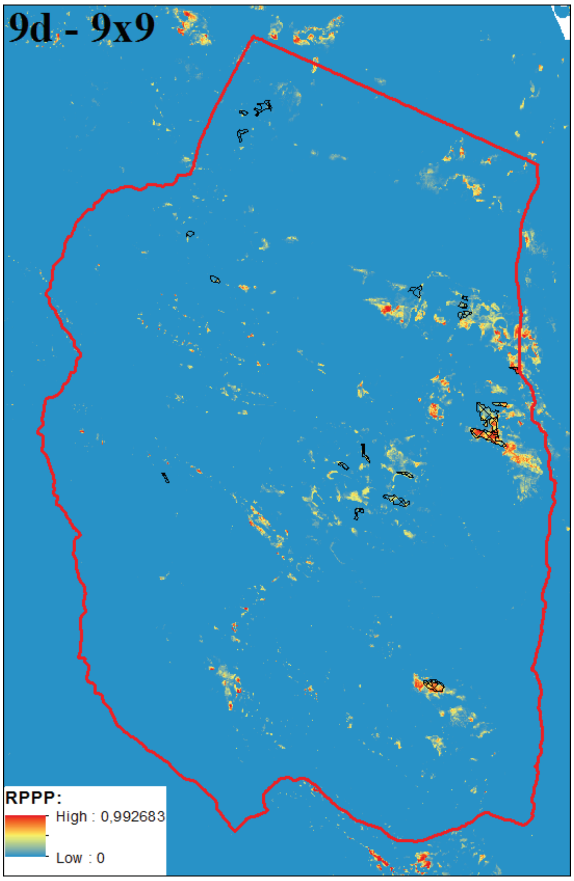

8d - 3x3
RPPP:
High : 0,964463
Low : 0


8d - 4.5x4.5
RPPP:
High : 0,985481
Low : 3,68155e-039


8d - 9x9
RPPP:
High : 0,930868
Low : 0
9c - 3x3
Value
High : 0,949759
Low : 9,08909e-012
9c - 4.5x4.5
Value
High : 0,885889
Low : 4,89071e-010
9c - 6x6
Value
High : 0,986805
Low : 5,26487e-029


9c - 7.5x7.5
Value
High : 0,994435
Low : 2,42481e-041
9c - 9x9
Value
High : 0,861306
Low : 1,76196e-007
9d - 3x3
RPPP:
High : 0,97714
Low : 0
9d - 4.5x4.5
RPPP:
High : 0,992683
Low : 0


9d - 7.5x7.5
RPPP:
High : 0,994997
Low : 6,95683e-040


9d - 9x9
RPPP:
High : 0,992683
Low : 0

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no