Centre for Geo-Information

Thesis Report GIRS-2006-22

Relating Alpine Treeline to Mountain Topography

A study in Sangay National Park, Ecuador

Johan Ruijten

May 2006

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Relating Alpine Treeline to Mountain Topography

A study in Sangay National Park, Ecuador

Johan Ruijten

Registration number 820604-719-090

Supervisors:

Maaike Bader
Arnold Bregt

A thesis submitted in partial fulfillment of the degree of Master of Science at Wageningen University and Research Centre, The Netherlands.

May, 2006
Wageningen, The Netherlands

Thesis code number: GRS-80436
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing
Thesis Report: GIRS-2006-22

Foreword

This MSc thesis is part of the PhD project “Spatio-temporal modeling of treeline dynamics” by ir. Maaike Bader of the ‘Laboratory of Geo-Information Science and Remote Sensing’. The topic allowed me to combine my interests in ecology with geographical information science. It also allowed to me rediscover the use (and even entertainment value…) of statistical analyses…

I have really enjoyed working on this topic the last few months and gained much experience during this research. Of course there are always things that I would have liked to do differently afterwards. But overall I must say that I’m quite satisfied with the result.

I would like to express my gratitude to Maaike Bader for her enthusiasm on the topic, helping me out during my work and her discussion on earlier drafts of this thesis. Special thanks to Arnold Bregt for his second opinion on the thesis and to Sytze the Bruin, John Stuiver and Jacob van Etten for their help and comments.

Index

List of tables

List of figures

Abstract

The aims of this study were to provide insight in the relationship between mountain topography and forest distribution at alpine treelines; to use this description to obtain a better insight in the processes and conditions influencing forest distribution at alpine treelines; and the design of a consistent method which could be used to compare forest distribution at alpine treelines between different areas.

A Landsat ETM+ image and an SRTM DEM of Sangay National Park, Ecuador, were used to derive forest distribution at treeline and several topographical and environmental indices. These variables were related to forest distribution by means of a logistic regression approach. The model used for prediction was cross-validated in the training area, and applied in a different study area nearby.

Forest distribution was mainly explained by altitude, topographic wetness (CTI), eastness and erosion potential (STCI). Predictive accuracy of the model ranged from 74.2 % to 84.0 % in the test and the training area. The test area has more human impact, which probably explains the over-estimation of forest there.

The logistic regression approach is suitable for discriminating the relative importance of the variables. The ecological meaning of some variables is hard to assess, because they affect several biophysical factors. The method developed during this research allows a quick investigation of factors with potential influence on alpine forest distribution, by using inexpensive and easy to obtain Landsat and DEM imagery. It also allows for a rapid comparison of forest distribution between different areas and the localization of potential disturbances.

1 Introduction

1.1 Alpine treeline

Treelines form a sudden transition from a habitat which sustains tree growth to a habitat which contains environmental constraints for tree growth. Throughout the world, treelines occur at a great variety of environmental gradients: e.g. thermal, drought, waterlogging, nutritional and salt stress gradients (Körner, 1998). The high altitude transition zone from montane forest (trees) to low alpine vegetation (grasses, shrubs, cushion plants) is defined as the alpine treeline ecotone, or ‘alpine treeline’ for short.

The alpine treeline is an intensively studied landscape boundary. An important reason for this is the potential sensitivity of the alpine treeline to climate change, induced by processes such as global warming. Several authors have investigated the relationship between shifting positions of alpine treelines and changing climate (Camarero & Gutiérrez, 2004; Daniels & Veblen, 2004; Dullinger *et al.*, 2004; Hansen *et al.*, 2001; Moen *et al.*, 2004). Alpine treelines could be useful as indicators of changing global climate, but there still remains much uncertainty about the (interacting) factors influencing tree growth at high altitude (Kupfer & Cairns, 1996). Moreover, most of these studies in alpine regions have focused on mid- and high-latitude mountain areas. Data from tropical regions are generally lacking (Hoch & Korner, 2005; Körner, 1998).

It is often assumed that one of the main limitations for tree growth at alpine treelines is available warmth (Daniels & Veblen, 2004; Körner, 1998), but there are also other factors used to explain variations in treeline position; for example soil properties, moisture, snow cover, geomorphic processes, and, not to forget, human influence (Allen & Walsh, 1996; Hoersch *et al.*, 2002; Körner, 1998; Stevens & Fox, 1991). Human influences such as anthropogenic burning and cattle grazing have a serious impact on treelines around the world. In some cases, natural, climatic treelines are even substituted by anthropogenic treelines. It is often difficult to discriminate between natural treelines and treelines which are limited in altitude due to human impact (Holtmeier & Broll, 2005; Kjällgren & Kullmann, 1998; Körner, 1998).

Insight is needed in the factors, influencing tree establishment at the uppermost limits of forest growth, to better understand the causes of alpine treeline and its spatio-temporal behaviour. Since there is a lack of research performed in tropical regions, this study focuses on providing insight in the factors influencing forest distribution at tropical alpine treelines.

1.2 Topography

Insight in the relation between forest distribution at alpine treeline and mountain topography will help to understand the processes and conditions that are important for treeline occurrence. Despite the large amount of research that has been performed on the influence of topography on vegetation status and tree establishment, there still remains a great deal of uncertainty about the influences of topographic factors on the spatial distribution of forest at alpine treeline position, especially in the tropics.

Topography is a factor that has been used by several authors to explain, predict or model vegetation growth near the alpine treeline border (Brown, 1994; delBarrio *et al.*, 1997; Dirnböck *et al.*, 2003; Hoersch *et al.*, 2002; Horsch, 2003; Paulsen & Korner, 2001). In most cases, the authors used a digital elevation model (DEM) to derive topographic variables. Hoersch et al. (2002) state an important reason for using a DEM instead of using spatial information that directly influences tree establishment: "*as spatial information on site factors is commonly lacking in mountain areas, the use of a DEM is a potential substitute for use in vegetation analyses, as it highly correlates with temperature, moisture, geomorphological processes and disturbance factors*".

First of all, this statement brings up the theory of topography being a factor having a strong indirect influence on vegetation growth because of its influence on parameters like temperature, available radiation, precipitation, air and soil temperature, soil moisture and nutrients, snow accumulation, wind etc. (Allen *et al.*, 1995; Bian & Walsh, 1993; Butler *et al.*, 1994; Kjällgren & Kullmann, 1998; Paulsen & Korner, 2001; Walsh *et al.*, 2003). A landscape with a spatially heterogeneous topography will therefore result in highly variable habitat conditions (Horsch, 2003).

Secondly, this statement brings forward the difficulty of obtaining environmental data in remote, mountainous areas. Direct factors such as climatic factors are usually measured as point data. To obtain a spatial distribution, this point data has to be interpolated. The resolution of factors derived from point data is therefore not very high. An alternative is to predict vegetation distribution by means of indirect factors (i.e. topographic variables) derived from a spatially continuous DEM (Guisan & Zimmermann, 2000).

Authors that used DEM-derived topography to predict or explain vegetation distribution at alpine treelines developed their methods mostly in temperate regions. In these regions vegetation is subject to seasonal differences and various topography-related factors not encountered in tropical, equatorial regions: e.g. snow accumulation and variation in solar irradiation due to north-south aspect (Brown, 1994; delBarrio *et al.*, 1997; Dirnböck *et al.*, 2002; Hoersch *et al.*, 2002; Horsch, 2003; Körner, 1998; Paulsen & Korner, 2001). We therefore expected to find different topography-treeline relationships in the tropics than have been found in temperate regions.

Factors that probably influence vegetation at tropical alpine treelines are for example soil temperature, drought and radiant cooling at night (Ohsawa, 1990). Little is known about the impact of topography on these factors. Tropical alpine treeline research could benefit greatly from a DEM-based approach for vegetation prediction since these regions generally have a shortage of high resolution field measurements of environmental factors.

Previous studies often used static, statistical models based on factor (PCA) analysis (e.g. Horsch, 2003), ordination techniques such as canonical correspondence analysis (e.g. Dirnböck *et al.*, 2003), discriminant analysis (e.g. delBarrio *et al.*, 1997) and logistic regression analysis (e.g. Virtanen *et al.*, 2004) to investigate the importance of topographical variables on vegetation distribution. The strength of these methods lies in the possibilities to explore individual relationships between the explanatory variables (topography) and the response variable (vegetation) (delBarrio *et al.*, 1997). However, in case of factor analysis (in which the set of explanatory variables is reduced to a substitute set of factors by rotating and translating the coordinate system) it is often very hard to interpret the ecological meaning of the factors (Garson, 1998; Guisan & Zimmermann, 2000; Horsch, 2003). Canonical correspondence analysis and discriminant analysis usually involves assumptions (such as multivariate normality) which are often not met when using topographical variables (delBarrio *et al.*, 1997; Garson, 1998; Press & Wilson, 1978).

Logistic regression has the advantage of having coefficients that are relatively easy to interpret which supports the assessment of ecological meaning (Garson, 1998). Furthermore, logistic regression avoids normality and linearity assumptions, which makes it a suitable method for dealing with variables that are not normally distributed or linearly related (Garson, 1998; Press & Wilson, 1978).

1.3 Aims & research questions

This study aims at evaluating the potential use of vegetation modeling at tropical alpine treelines based on vegetation-topography relationships at landscape level. Hence, the objectives of this study are to:

1) Provide insight in the relationship between forest distribution at alpine treelines and mountain topography.
2) Use this description to obtain a better insight in the processes and conditions influencing forest distribution at alpine treelines.
3) Design a consistent method which can be used to compare forest distribution at alpine treelines between different areas.

The above research objectives result in several concrete research questions divided in ecological and methodological questions:

Ecological research questions

a) What are the effects and relative importance of topography and topography-related environmental factors on the spatial distribution of forest at alpine treeline?
b) What part of forest distribution at alpine treeline can be explained by topography and topography-related environmental factors?

Methodological research questions

c) Which topographical and topography-related environmental variables can be used in order to reach the first objective and how will they be derived?
d) How can the relationship between these variables and forest distribution at alpine treeline be quantified?
e) How does the method perform in a different area with a similar environment and how does it perform in an area with a different environment; i.e. how do the changed conditions affect the spatial distribution of forest?

First, the statistical relationships between DEM-derived variables and forest distribution are explored. We use a logistic regression approach to build a model to explore the relationships between the topographic variables and to predict forest distribution at the alpine treeline in Ecuador.
Next, the model derived from the statistical relationships is used in a different part of the study area where human impact is more intense, in order to predict potential forest distribution.

The next chapter provides more detail about the material and methods used in this research. After that, the results are presented. We debate our findings in the chapter discussion and finish this report with the most important conclusions and some recommendations for further research.

2 Material & Methods

2.1 Study Area

The study area is located in the central Andes mountain region, about 20 km south-east of the city of Riobamba, Ecuador. It is part of Sangay National Park which is located between the three provinces Tungurahua, Chimborazo and Morono Santiago. This national park comprises approximately 540.000 ha. (Armstrong & Macey, 1979)

Within the perimeters of the park, three main geomorphologic zones can be found: volcanic high Andes, eastern foothills and alluvial fans. The alpine treeline is located in the high Andes zone, which is characterized by a very heterogeneous landscape with steep valleys and high peaks. The main rivers are draining rapidly, and with high erosive power, to the east into the Amazon basin (UNEP-WCMC, 2005).

The climate ranges from subtropical to temperate. The eastern slopes receive most rainfall (extremes of 5000 mm/year have been recorded) while the western slopes receive not more than 600 mm/year. The snowline occurs at 4800 m. (Armstrong & Macey, 1979; UNEP-WCMC, 2005).

Figure 1: Vegetation types in the treeline transition zone; on the left we can see montane cloud forest, on the right tussock grasses

In the transition zone from forest to non-forest a number of vegetation types can be distinguished. Areas located above treeline altitude are mainly dominated by low growing species including grasses - such as *Calamagrostis sp.*, *Festuca sp.* and *Stipa sp.* - and various shrubs and cushion plants (see figure 1). These grassland areas are also called ‘páramos’. They are found down to an altitude of 3700 meters (Armstrong & Macey, 1979).

Montane cloud forests are found approximately below 3750 meters (figure 2). Trees in the upper regions of this area are approximately 5 meters high and are dominated by species such as *Escallonia myrtilloides* and *Gynoxys buxifolia*. Trees in the lower regions reach heights of 12 meters. In these regions there is a greater variety of species, like *Senecio vaccinoides, Diplostephium, Brachyotum, Hesperomeles, Buddleja*, and *Miconia* (Armstrong & Macey, 1979; UNEP-WCMC, 2005).

Figure 2: high montane cloud forest in Sangay National Park. The transition from small trees to paramo grasslands is clearly visible.

Inside Sangay National Park, two smaller areas were chosen as study area. One area is located North of Sangay volcano and is approximately 25 x 25 kilometers in dimension (north-west corner: 78º29' W, 1º46' S). This ‘*training area*’ was used to develop the model and to obtain knowledge about the factors influencing forest distribution inside the alpine treeline zone. The second area is approximately 30 x 30 km in dimension and is located more to the south (north-west corner: 78º35' W, 2º4' S). This '*test area'* was used to apply the model and to test its performance (see figure 3).

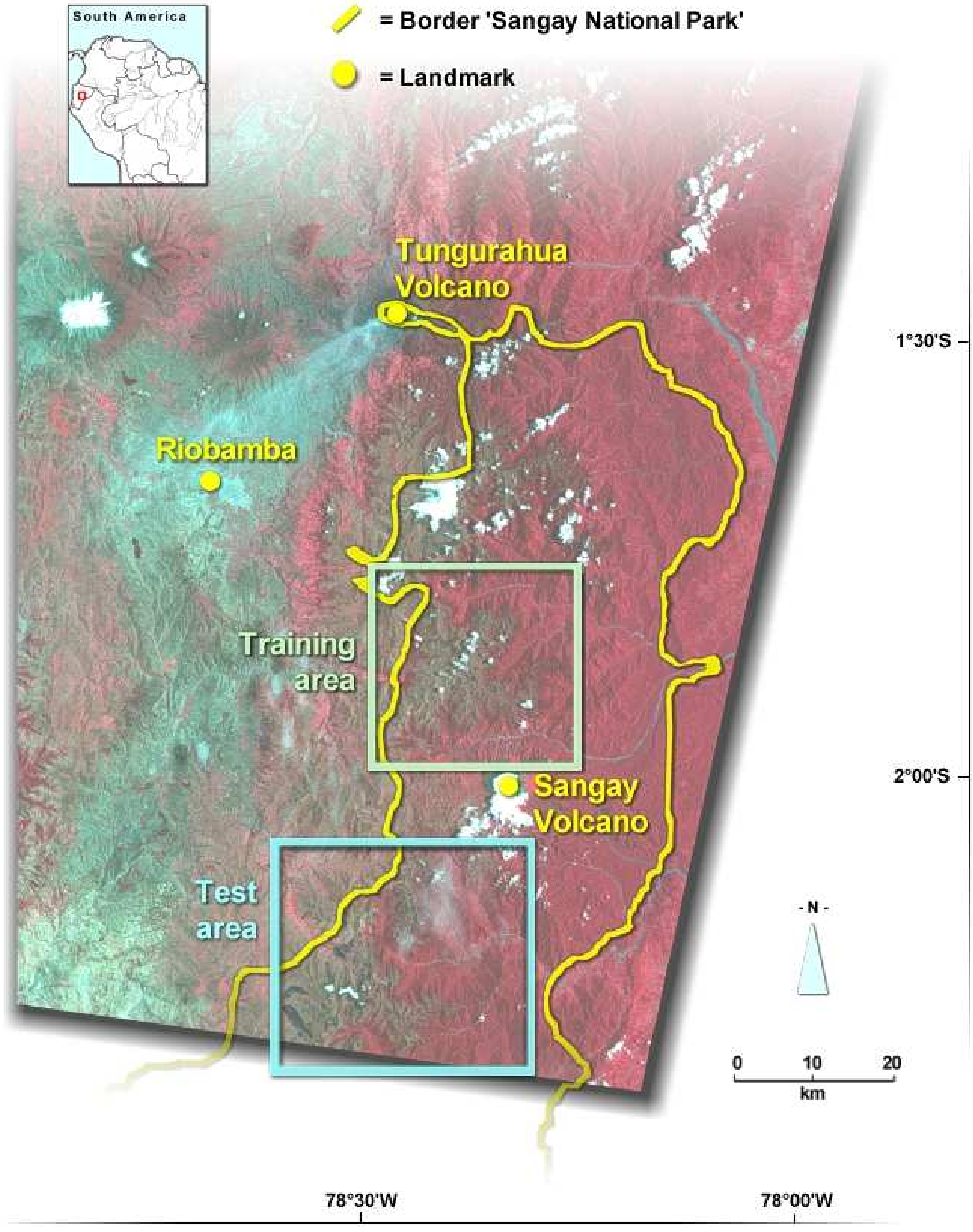


Figure 3: Map of Sangay National Park, Ecuador.
The map shows the location of the 'training area' and the 'test area' inside Sangay National Park. The 'training area' is approximately 25 x 25 kilometers in dimension (north-west corner: 78º29' W, 1º46' S). The 'test area' is approximately 30 x 30 km in dimension (north-west corner: 78º35' W, 2º4' S).
The map is drawn upon the Landsat image displayed in false colours (red = band 4, green = band 3, blue = band 2).

2.2 Source Data

The source data consisted of an ortho-rectified Landsat ETM+ image (in GeoTIFF format) and a Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) (in GeoTIFF format). The Landsat image was acquired in September 2001 and has a spatial resolution of 28.5 meters and a horizontal accuracy of approximately 64 meters. The DEM was obtained in 2000 by the USA Global Land Cover Facility. It has a spatial resolution of 90 meters and a horizontal and vertical relative accuracy of approximately 15 respectively 6 meters. The data was already geo-referenced by the supplier. In case of the DEM the values were also rounded to integers, resulting in 1 meter intervals (GLCF, 2006; Rodriguez *et al.*, 2005; USGS, 2003)

The Landsat image was used for identification of the forest area and extraction of the alpine treeline position. The DEM was used to derive several topographic variables and environmental indices. See figure 4 on page 21 for a general overview of the data and processes used in this research. The following text explains this dataflow in more detail.

2.3 Preprocessing

Before the source data was suitable for use in this research, a number of preprocessing techniques were applied.

First, the training area was chosen based on the following criteria: it had to contain a clear transition from forest area to non-forest area, and the amount of pixels with missing values in the DEM and in the Landsat image (due to cloud cover) had to be as small as possible. In this case, the resulting area was located around 78º29' W, 1º46' S (north-west corner) and was about 25 x 25 kilometers in size. The separate Landsat bands (bands 1-5 and 7) were then stacked into a composite image file. Together with the DEM, they were clipped to the selected area.

The DEM image contained some missing values over land and negative values in water bodies. These bad pixels are often found in radar shadow of DEM's created by using the interferometric radar technique (GLCF, 2006). To overcome this problem a surface fitting technique, such as the Delauney triangulation method, can be applied. This method uses the values of surrounding pixels to fill the 'gap' with triangles (Moore *et al.*, 1991). This method is available for example in the software package ENVI from RSI. The function is called 'Replacing bad values', whereby the user defines which value or range of values are considered 'bad' and are to be replaced by calculating new values using Delauney triangulation.

Another problem with the DEM was that it contained 'terraces', which showed as diagonal lines of flat areas throughout the image when for example calculating slope angle. This is probably caused by the process of geo-referencing the source data to the WGS84 ellipsoid with elevation data rounded to integer values (Wood, 2003). There are a few methods to deal with this problem. The best option would be that the provider would scale the z values by 10 or 100 before applying the interpolation procedure, thereby removing the errors of rounding values with a meter interval (Wood, 2003). However, when this option is out of range because the available data have already been interpolated, some alternative methods are available. Two methods were investigated of which the second one was selected for this research.

The first method uses a mean filter which passes over the raster image. It is a so called ‘focal’ filter. For every pixel in the image, the mean value of its surrounding pixels is calculated using a matrix (or kernel) often with a size of 3 x 3 pixels. The resulting mean value is then assigned to the central pixel. This method smoothes the image by reducing the variation between one pixel and its neighboring pixels. To remove the bad lines in this particular case, a matrix size of 5 x 5 pixels is required. The function is available in ESRI’s ArcGIS spatial analyst extension.

The second method transforms the raster image to an image build up of elevation contour lines. Then, using the ‘Topo to raster’ tool available in ArcGIS spatial analyst, the contour data is converted back to a raster image. The interpolation method used in this tool is developed by Hutchinson and is based on the ANUDEM program developed in 1988, 1989 (Hutchinson, 1989). It interpolates elevation values following an iterative approach by taking into account that the resulting DEM must have a connected drainage structure. In this particular case, the contour lines were created with a 15 meter elevation interval. For the conversion from contour to raster, the output cell size is set to 90 meters, 50 interpolation iterations are used, and drainage enforcement is put on.

Both methods presented above smooth the original dataset and thereby loose some basic elevation data, but both methods effectively remove the terraces. The last method has the advantage that it creates a hydrological correct elevation model by removing sink points from the input data. This is useful when using the DEM to calculate environmental indices later on. Therefore, the latter method was selected to reduce the striping error in the original DEM.

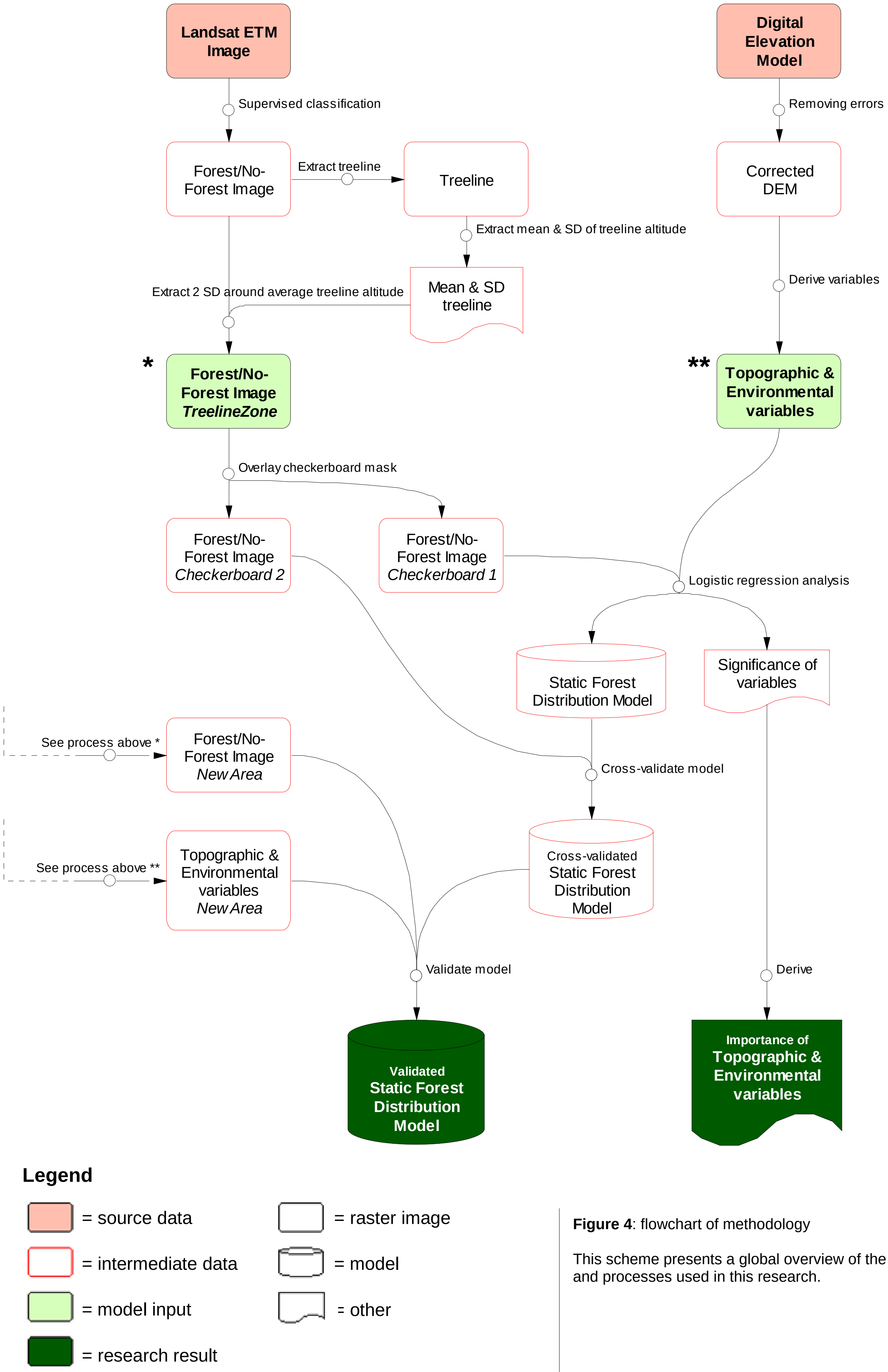


Figure 4: flowchart of methodology

This scheme presents a global overview of the data and processes used in this research.

2.4 Classification

When dealing with a Landsat image that suffers from heavy topographic shading it is recommended to first apply a simple correction method such as a cosine correction:

$$CorrectedRadiance = RawRadiance*(\cos \theta / \cos i)$$

In which θ is solar zenith angle at the time of acquisition and i is the local incidence angle, which can be calculated by using the DEM and the equation below:

$$\text{Cos } i = \cos \beta * \cos \theta + \sin \beta * \cos \theta * \cos (\lambda - \Phi)$$

In which β is the local slope angle in degrees, Φ is the solar azimuth angle at the time of acquisition and λ is the local aspect (Riano *et al.*, 2003; Shepherd & Dymond, 2003). In this case, the topographic shading did not interfere with the classification of forested areas and was therefore not performed.

In order to extract the treeline from the Landsat image, the areas of forest and non-forest had to be correctly identified. For this purpose, two methods were investigated.

The first method is based on the calculation of a vegetation index, such as the Normalized Difference Vegetation Index or NDVI. It is available in many GIS software packages, in RSI Envi it is available under 'Transform > NDVI'. This index uses band 4 (near infrared) and band 3 (red) to calculate the probability that a pixel represents vegetation. The result is an image with pixel values ranging from 0 to 1 in which values closer to 1 are more likely to represent vegetation. Next, by reclassifying values bigger than a certain threshold as forest, a forest/non-forest image is created.

The second method is to perform a supervised classification. The image was first displayed as a false color composite (band 4 – red, band 3 – green, band 2 – blue) and the original band values were linearly stretched over a range from 0 to 255 for a clear visualization. Then, by selecting a number of training sites in areas that were known to be covered by forest, the spectral signatures of these locations were identified. These signatures could then be used to identify the cover of the remaining part of the image. In this particular case a maximum likelihood decision rule with a probability threshold of 0.995 was chosen.

Both methods presented above gave similar results. Because of a lack of ground truth data, the result of both methods can only be analyzed visually by comparing it to forest cover in the original Landsat image, which is well recognizable. In this research, the result of the supervised classification method yielded the most satisfactory result, based on the visual comparison with the Landsat image, and is therefore used in further analyses.

Finally, we assumed a mainly closed forest area. Therefore we removed individual groups of forest pixels by using the 'sieve'-tool in IDL Envi with a group threshold of 4 pixels. Small gaps in the forest area are removed by using the 'clump'-tool in Envi with a group threshold of 2 pixels.

2.5 Treeline zone

The following procedures were all carried out in ESRI ArcGIS 9. First, to be able to extract the treeline, the classified raster image was converted to a 'coverage'. In this way, the boundaries between forest and no forest were represented as lines. For further analysis, these lines were converted back to a raster image with the same cell size as the original Landsat image (28.5 meters). During this process, a mask, which represented the clouded areas in the Landsat image and the outer boundaries of the study area, was used to remove these undesired areas from the treeline image.

In the resulting image every boundary from forest to non-forest is visible, even if they do not represent the alpine/altitudinal treeline. However, only the alpine treeline should be used to calculate the average treeline position and its standard deviation. These were necessary to define the treeline transition zone within which the treeline position was modeled. Therefore a minimum alpine treeline height is set. Forest boundaries below this height were excluded from the treeline image.

The treeline transition zone ('treeline zone' for short) was defined as the area within two standard deviations from the average altitude of the alpine treeline. To determine the average altitude, the treeline raster image was used as an extraction mask on the DEM. The result was a treeline image in which every pixel has an elevation value. The mean and standard deviation of these elevation values can be read out of the metadata of the image or put into a table by calculation of the image statistics. The average treeline position was located at 3634 meters with a standard deviation of 185 meters. The transition zone was located between 3264 and 4004 meters elevation.

From the DEM, the pixels which have a value between these limits are extracted. This image is then used as an extraction area on the forest/non-forest image. The result is an image representing forest distribution in the alpine treeline transition zone.

2.6 Topographic variables

From the DEM, the following topographic variables were calculated using ESRI's ArcGIS 9 spatial analyst extension: slope angle, aspect, plan curvature and profile curvature. These variables are calculated using a 3x3 moving window. Since aspect is expressed in degrees from 0 to 360, low values are actually the same as high values. Therefore, two other aspect values are calculated. By applying a cosine transformation a variable is created stressing the north-south contrast. A sine transformation produces an image stressing the west-east contrast. Values range from 1 to -1, where 1 represents northern, respectively eastern aspects and -1 represents southern, respectively western aspects. The drawback of this method is that the new sine and cosine values change with a variable amount i.e. the intervals are not constant, which is an assumption in a number of statistical analyses (Jenness, 2005). However, the statistical method presented later in this research does not make this assumption.

Finally, for further analysis, the topographic raster images need the same spatial resolution as the forest distribution image. Therefore, the topographic raster images are resampled from a cell size of 90 meters to 28.5 meters using a bilinear interpolation process. Bilinear interpolation uses a distance weighted average of the values of the four nearest pixels. Since this resampling method uses the relative position of its neighboring pixels, bilinear interpolation is preferred for data which represents continuous surfaces such as elevation, slope etc. (ESRI, 2001).

2.7 Environmental indices

Since topographic variables are possibly too limited in explaining the distribution of forest at treeline position, we used a number of indices for environmental conditions which may have a direct influence on vegetation growth. The indices were derived from the DEM, which restricts the number of options. The chosen indices represent conditions which are known to influence vegetation growth and/or establishment: solar radiation, wetness and erosion potential.

2.7.1 PRR

The first index, the ‘Potential Relative Radiation’ index, (Pierce *et al.*, 2005) represents incoming radiation during the year. High values indicate high potential incoming radiation. The calculation of this index accounts for daily and annual changes in solar orientation as well as topographic shading effects. Since aspect and slope fail to capture temporal changes in incoming radiation as well as the effect of topographic shading, this index might prove to be a valuable addition to the set of explanatory variables.

Calculation of PRR is as follows: first calculate solar elevation (complement of solar zenith) and solar azimuth in degrees for every hour from sunrise to sunset. Calculate these values for the day of the month which represents the average solar period of that month. To calculate the solar elevation and solar azimuth per hour of the day for a particular geographic location, the ‘Solpos’ algorithm, developed by the National Renewable Energy Laboratory (NREL), was used (NREL, 2001).

When solar elevation and azimuth values had been obtained, this data was then used together with the DEM to calculate hourly shaded relief grids. This was

done with the HILLSHADE function in ArcInfo. By turning on the shadow option in this algorithm, not only local shade but also shadowing effects of nearby hills were taken into account (ESRI, 2001).
Next, the hourly values were summed. This resulted in daily totals which represented monthly averages. Then, these monthly averages were summed which resulted in a yearly map of potential incoming radiation.

2.7.2 *CTI*

The second index, the Compound Topographic Index, commonly referred to as Wetness Index, represents soil water content and indicates zones of water saturation. High values indicate converging, low areas. Low values indicate diverging, steep or high areas (Schmidt & Persson, 2003; Yang *et al.*, 2005). It is a static wetness index and is most commonly calculated in two different forms:

1) $CTI_T = \ln(A_s / T \tan\beta)$ or,
2) $CTI = \ln(A_s / \tan\beta)$

In which As is the upslope contributing area (or catchment area in area (m^2) per unit width orthogonal to flow direction), β is the local slope angle and T is the soil transmittance when it is completely saturated (Gessler *et al.*, 2000; Moore *et al.*, 1991). Since T is not available for the study area, CTI is calculated by using the second form. This form assumes uniform soil properties, which is a limitation but earlier research showed strong correlation of CTI 2 with several soil attributes such as surface soil water content, horizon depth, silt percentage, organic matter content and phosphorus (Gessler *et al.*, 2000; Moore *et al.*, 1991; Yang *et al.*, 2005).
The algorithm D∞ ($D_{infinity}$) as proposed by (Tarboton, 1997) is used to acquire A_s which can be found in the program TauDEM (Terrain Analysis Using Digital Elevation Models). TauDEM is available as an extension for ESRI's ArcMAP software. The algorithm has the advantage that it is proportioning flow between two downslope pixels, thereby overcoming the problems of parallel flow lines (encountered with the basic D8 algorithm (Schmidt & Persson, 2003; Yang *et al.*, 2005)) and unrealistic dispersion (encountered with multiple flow algorithms (Freeman, 1991; Quinn *et al.*, 1995)) (Tarboton, 1997). TauDEM was used to calculate flow direction which, in turn, was used to derive the upslope contributing area. The remaining calculations were automated using ESRI's ArcGIS modelbuilder.

2.7.3 STCI

The last index, the Sediment Transport Capacity Index (STCI), is a measure for erosion potential. It is an index that is suitable for determining erosive power at the scale of catchments and has proven successful to predict soil properties which influence vegetation distribution (De Roo, 1998; Dirnböck *et al.*, 2002; Moore *et al.*, 1993). The calculation of STCI follows on the Universal Soil Loss equation (USLE) and is similar to the Stream Power Index (SPI). It was developed by (Moore *et al.*, 1993). STCI is expressed as:

$$\text{STCI} = [A_s/22.13]^{0.6} / [\sin \beta/0.0896]^{1.3}$$

In which A_s is the upslope contributing area (m^2/m) and β is the slope angle (degrees) (Moore *et al.*, 1993).

The choice for STCI instead of USLE and SPI is based on the consideration that the expression contains the upslope contributing area (A_s). Therefore, the index accounts for the convergence and divergence of flow. In this way, the index is more suitable for a landscape with a complex topography as in case of this research (Moore *et al.*, 1991).

The calculation of A_s was performed by using the same D∞ algorithm as for CTI. The remaining calculations are automated using ESRI's ArcGIS modelbuilder.

2.8 Statistical Analysis

2.8.1 Development of the logistic regression model

For the study area we now had the following maps at our disposal: the dependant variable forest/no-forest, and the independent variables height, aspectNS (north-south), aspectEW (east-west), slope angle, plan curvature, profile curvature, CTI (wetness), PRR (radiation) and STCI (erosion). The objective was to explain the variable forest/no-forest by means of the other variables i.e. the occurrence of forest is predicted by a set of explanatory variables.

Since the variable forest/no-forest is a binomial or dichotomous variable, binomial logistic regression seems the most obvious statistical method to analyze the dataset (Garson, 1998; Hosmer & Lemeshow, 1989). Logistic regression is widely used in ecological researches dealing with a binomial or multinomial dependant variable explained by independent variables of any type, also in the field of vegetation prediction (Augustin *et al.*, 2001; Augustin *et al.*, 1996; Calef *et al.*, 2005; Felicisimo *et al.*, 2002; Hilbert & Ostendorf, 2001; Virtanen *et al.*, 2004). Logistic regression does not require the independent variables to be in an interval scale which is useful when including the sine and cosine transformed variables aspectNS and aspectEW into the analysis (Garson, 1998).

In logistic regression, the probability of forest occurence can be expressed as a function of the explanatory variables:

$$P(\text{forest}) = \frac{e^{(\text{constant}+c_1\cdot\text{variable}1+c_2\cdot\text{variable}2+\ldots+c_n\cdot\text{variable}n)}}{1+e^{(\text{constant}+c_1\cdot\text{variable}1+c_2\cdot\text{variable}2+\ldots+c_n\cdot\text{variable}n)}}$$

In which P is the probability on the occurrence of forest and c_n represent coefficients related to the explanatory variables (Felicisimo *et al.*, 2002; Hosmer & Lemeshow, 1989).

The logistic model was developed in the '*training area'* following a checkerboard pattern (see figure 5 below). The red areas inside the alpine treeline zone were used for the development of the logistic model ('*development area*'). The green areas (*'cross-validation area'*) were used to cross-validate this model. The model was also applied in a new area approximately 30 kilometers to the south of the development area. This *'test area'* is located close to the border of the national park. Therefore, this area is somewhat different compared to the development area, especially in terms of land use. Parts of the area are used by people to feed their cattle. For validation purposes this area is therefore not very appropriate. However, the results can be used to locate areas where forest could potentially occur but is absent due to human interference.

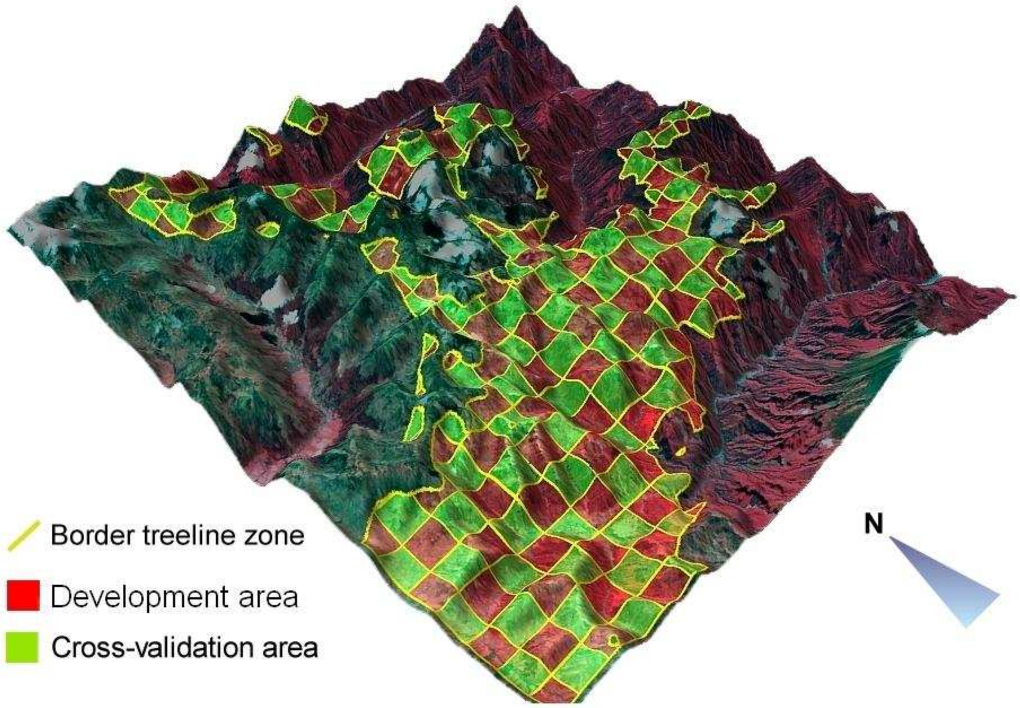


Figure 5: checkerboard pattern representing the *'development area'* and '*cross-validation area*'
The treeline zone was dived into two parts following a checkerboard pattern. The red zones were used for the logistic regression analysis and development of the model. The green zones were used for cross-validation of the model.

(scale varies in this perspective, distance from north to south is approximately 25 km)

To apply the logistic regression analysis, the data was first transferred to the statistical package SPSS. This was done by transforming the raster images to point features. In this way the value of every pixel receives a point-id which is useful for visualization of the predictions later on. By extracting the tables of the point features, the values of the variables could be read into the SPSS software.

In SPSS the ‘binary logistic’ regression method was chosen with the variables entered into the model following a forward stepwise algorithm based on a classical maximum likelihood ratio test (StepwiseLR). The inclusion of variables is based on Rao’s score statistic, which is related to a likelihood ratio test of the coefficient of an explanatory variable. However, compared to a likelihood ratio test, the score statistic is computationally much faster because it is a non-iterative method (Garson, 1998). For excluding variables from the model the algorithm uses the change a variable causes in the -2 Log Likelihood (i.e. goodness of fit of the model). If a variable in the model causes no significant change in the –2LL it is removed from the model. The model coefficients are iteratively adapted to make the likelihood of the observed data as large as possible. This stepwise procedure allows non-significant variables to be excluded from the model. (Garson, 1998; SPSS, 2003; Wuensch, 2005).

Following this method we assume that the data points are independent from each other. However, this is not true because these points are spatially structured and are likely to show spatial autocorrelation. Ecological studies have sometimes included spatial autocorrelation into their model by means of autocorrelative models (e.g. Augustin *et al.*, 1996; Brenning, 2005). However, including spatial autocorrelation greatly complicates the application of such a model for predictions, so it should only be included when dealing with patchiness caused by factors other than biophysical factors (Augustin *et al.*, 1996). In this study we assume that all spatial pattern is caused by topographical factors, and not by purely spatial processes. At the current scale of investigation it is not likely for a forest pixel to be influenced by the vegetation cover of neighboring pixels. Most neighbor influence will occur at the scale of individual trees e.g. local seed dispersion, shelter. Measures to account for spatial autocorrelation are therefore not included in our model. However, the spatial dependence of the data does affect the statistical significance of the model, which can therefore not be interpreted directly (Guisan & Zimmermann, 2000).

When the explanatory variables are known, the equation can be used to estimate the probability on the occurrence of forest in the ‘*cross-validation area*’ and the ‘*test area*’. Probability values (P) are distributed between 0 and 1. A threshold value can be used to define the state of the response variable *forest/no-forest*. By default, probability values under 0.5 are classified as no-forest, and values above 0.5 are classified as forest.

2.8.2 *Importance of variables*

There are a few possibilities to assess the importance of the variables in explaining the variable forest/no-forest. First of all, we used the change that a variable causes in the -2LL when it is removed from the final model. A high change indicates high importance (Garson, 1998).

We compared this statistic with the outcome of the Wald statistic, which is the square root of the ratio between the model coefficient and its associated Standard Error (S.E.) (Garson, 1998; Hosmer & Lemeshow, 1989; Wuensch, 2005). Here also, a high value indicates high importance.

If the Wald statistic contradicts the outcome of the -2LL test, then the -2LL test is chosen as assessment statistic, since the Wald statistic has been criticized for being unstable and lacking sufficient statistical power (Garson, 1998; Hosmer & Lemeshow, 1989; Wuensch, 2005). We did assess the Wald statistic since it is useful to have an alternative statistic to support the outcome of the -2LL test.

For an even better perspective on the outcome of both the -2LL-method and the Wald statistic we assessed the actual predictive capabilities of the model in each step. It is important to note, however, that this can not be used as an assessment of the goodness of fit of the model, since it does not use the actual predicted probabilities but the cut-off values 0 and 1. Therefore, the outcome of the predictive capability test gives no insight in how far the predicted probabilities are located from 0 or 1 (Garson, 1998).

All variables that had a significant change on the -2LL were included in our '*full model*', since in our case this produced a model with the highest predictive capability when using a stepwise algorithm. However, to assess the importance of the individual variables in explaining forest/no-forest we had to beware of correlation between these variables. Since we dealt with variables that were in one way or another derived from each other, the risk on multicollinearity is high. Multicollinearity is defined as the intercorrelation between the independent variables (Garson, 1998; Hosmer & Lemeshow, 1989). High intercorrelation makes it very difficult to assess the relative importance of the correlated variables. The non-correlated variables are not affected (Garson, 1998).

In order to find out if the independent variables in the final model displayed signs of multicollinearity, we evaluated the bivariate Pearson correlations between the independent variables in the final model. We used a correlation value of 0.9 as threshold for indicating signs of multicollinearity (Garson, 1998).

However, the correlation matrix only displays signs of bivariate multicollinearity, but our model contained more than two independent variables. Therefore we produced multicollinearity diagnostic statistics to assess multivariate multicollinearity (Garson, 1998; Guisan & Zimmermann, 2000). These statistics produce the Tolerance factor and the Variance Inflation Factor (VIF) which is 1/Tolerance. VIF is defined as "*the number of times the variance* *of the corresponding parameter estimate* (= variable coefficient) *is increased due to multicollinearity as compared to as it would be if there were no multicollinearity*" (SSTARS, 2005). In case of logistic regression, values above 2.5 may indicate multicollinearity (SSTARS, 2005).

Independent variables showing signs of multicollinearity were separately entered in a new stepwise logistic regression analysis excluding the correlated variables, thereby creating stripped models that allow better assessment of the individual importance of these variables in explaining forest distribution in the alpine treeline zone. To compare the goodness of fit between models we used the likelihood ratio test. Since -2LL follows a chi-square distribution the outcome of this test is a chi-square statistic (Garson, 1998).

3 Results

3.1 Model Input

The figures below show the results of the preprocessing methods applied to respectively the Landsat ETM+ image and the DEM. The result of the classification (figure 6 below) is a variable representing forest and no-forest in the alpine treeline zone (3264 - 4004 meters). The figure shows a largely continuous forest area with a clear transition from forest to non-forest. However, it can be also seen that in various places the continuous forest area is interrupted by some small gaps or larger strips of non-forest area.

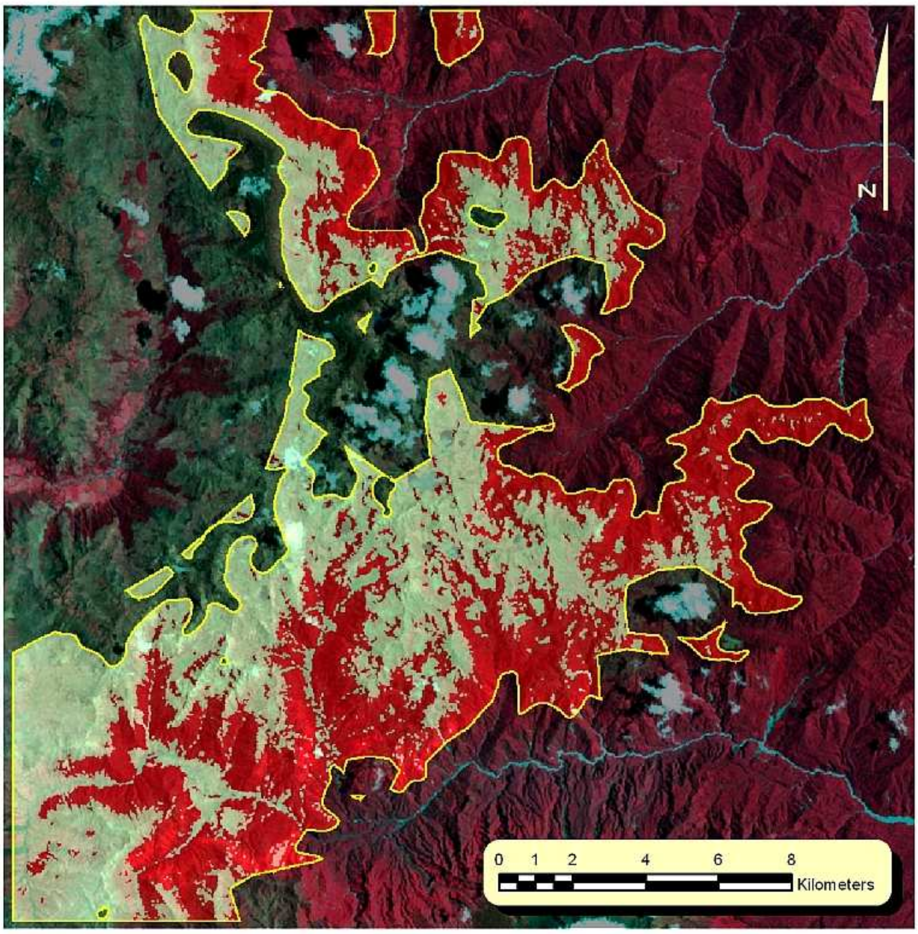
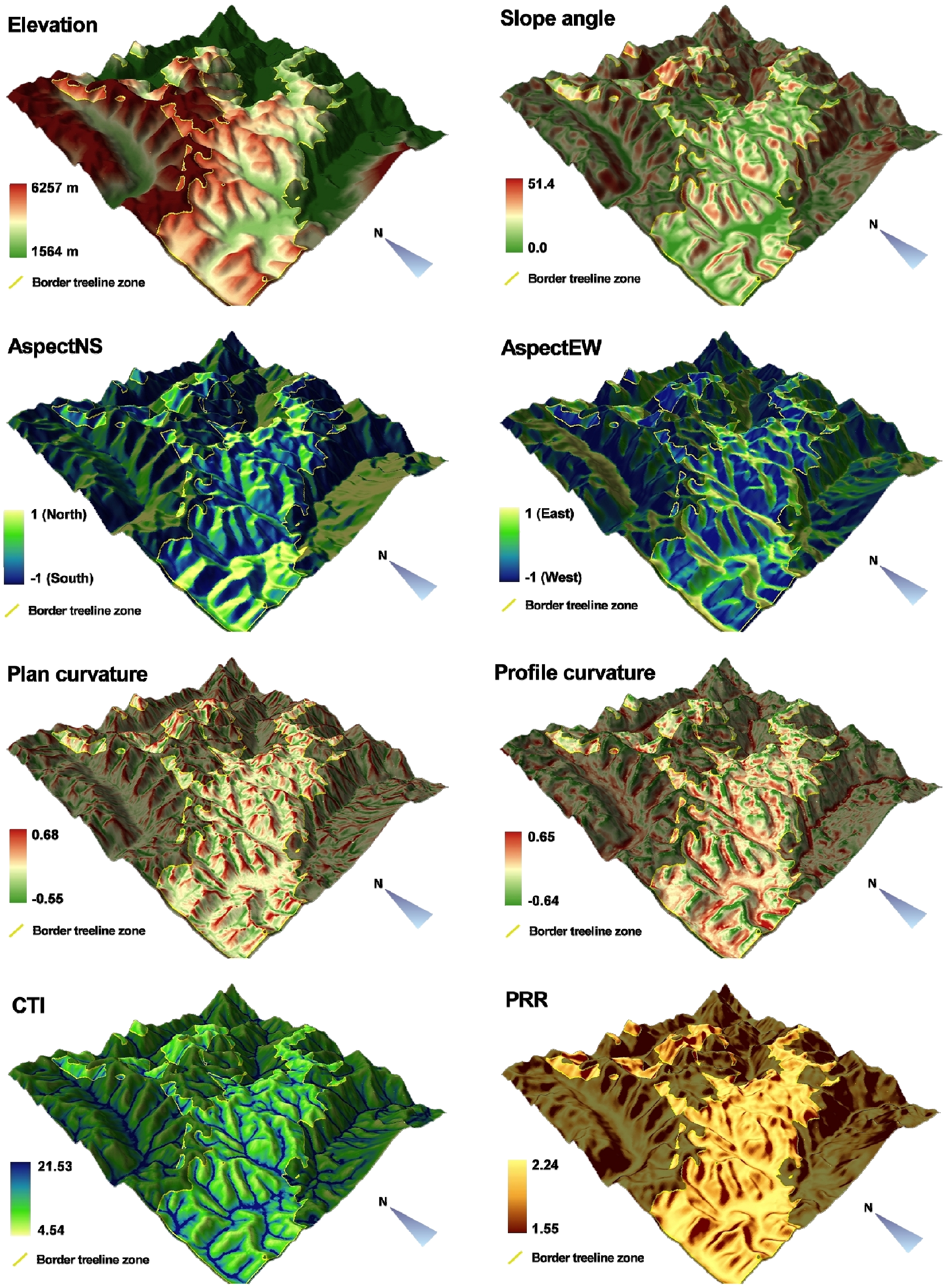


Figure 6: classification result of Landsat ETM+ image.
The Landsat image is displayed as a false colour composite (red = band 4, green = band 3, blue = band 2). The highlighted area with yellow boundaries indicates the alpine treeline zone. Inside this zone the red area represents the classified forest area. The greenish area represents no-forest.

Figure 7 shows the various topographic and environmental variables derived from the DEM. For a better comprehension, these variables are presented in a 3 dimensional perspective. While some variables have a clear unit of measure (height in meters, slope in degrees), other variables such as Aspect, PRR, etc., are displayed in relative units.



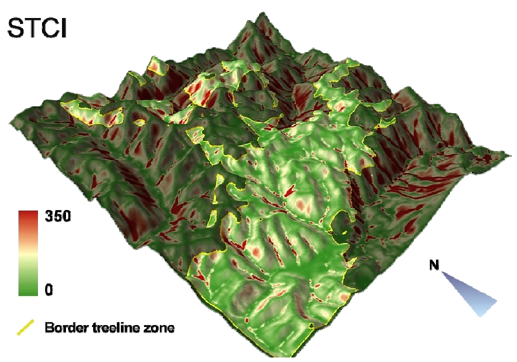


Figure 7

Topographic and environmental variables; from top-left to lower-right corner: **elevation, slope angle, aspect (north-south), aspect (east-west), plan curvature, profile curvature, CTI (wetness), PRR (radiation), STCI (erosion)**
The highlighted area indicates the treeline zone

Scale varies in this perspective; distance from north to south is 25 kilometers

When looking at the wetness index (CTI), a clear structure is emerging showing low potential water accumulation on ridges and high potential accumulation in valley bottoms and gullies. When looking at the potential radiation index (PRR), we can see that the highest potential is located on ridges and in valley bottoms and gullies. Areas with high potential erosion (STCI) are mainly located on the steeper slopes and in areas with a high upslope contributing area.

3.2 Logistic Regression Analysis

3.2.1 Full model

The variables described above were used as inputs in a stepwise logistic regression analysis. The results of the final step (the full model), as well as a summary of the variables entered in previous steps, are displayed in table 1. The variables are ordered from high to low influence according to the Wald statistic, which will be described in more detail later on. The total model chi-square is the measure of the goodness-of-fit of the model. It is used later on to compare the stripped models with this full model. A higher chi-square indicates a better fit.

Table 1: Variables in the 'Full model'.
The table presents the results of the final step (step 8) of the stepwise logistic regression algorithm. The variables are ordered from high to low influence according to the Wald statistic. Model Coefficient (B) shows the predicted effect of the variables on the odds ratio forest/no-forest. Positive coefficients represent a positive effect, and vice versa. S.E. shows the associated standard error. Exp(B) represents the odds ratio. An Exp(B) >1 indicates increased odds of forest occurrence. Sig. shows the significance of the contribution of the variables to the model. Total model Chi-Square is a measure of the goodness-of-fit of the model. Higher Chi-Square indicates better fit.

Step	Variable	Model Coefficient (B)	S.E.	Wald	Sig.	Exp(B)
Step 8	**Height**	-0.01311	0.00007	**33185.48**	0.00	0.98697
Full Model	**AspectEW**	-0.75409	0.01332	**3207.19**	0.00	0.47044
	CTI	-0.50856	0.00978	**2701.91**	0.00	0.60136
	Slope	0.06243	0.00281	494.70	0.00	1.06442
	AspectNS	-0.24278	0.01311	343.06	0.00	0.78444
	STCI	0.00701	0.00041	290.06	0.00	1.00703
	PRR	3.19912	0.20847	235.49	0.00	24.51099
	Plan curvature	-0.95960	0.10560	82.57	0.00	0.38305
	Constant	42.69393	0.56869	5636.22	0.00	3.48127E+18

Variable entered on step 1: Height
Variable entered on step 2: CTI
Variable entered on step 3: AspectEW
Variable entered on step 4: Slope
Variable entered on step 5: STCI
Variable entered on step 6: AspectNS
Variable entered on step 7: PRR
Variable entered on step 8: Plan

Variable excluded of model equation: Profile curvature

Total Model Chi-Square: **79698.95**

The column ‘Model Coefficient (B)’ shows the predicted effect of the variable on the odds ratio of forest/no forest. Positive coefficients represent a positive effect, while negative coefficients indicate a negative effect.

Column ‘Exp(B)’ is the actual odds ratio; if Exp(B) is >1, a higher value of the variable corresponds to increased odds of forest occurrence, if Exp(B) is <1 it corresponds to decreased odds of forest occurrence (Garson, 1998; Hosmer & Lemeshow, 1989; SPSS, 2003; Wuensch, 2005). Exp(B) of the variable ‘height’, for example, is <1 which implies a decrease in forest probability with increasing height. For aspect (east-west) it is also <1, which indicates a decrease in forest probability on eastern slopes; i.e. forest seems to prefer western slopes.

Profile curvature is not included in the model because of its low significance; 0.753 with an exclusion value of 0.05. See figure 8 below for an overview of possible curvatures. It is important to note that the outcome of the curvatures calculated with the algorithm implemented in ArcGIS deviates from normal: negative curvature normally indicates concave areas, positive curvature convex areas. Plan curvature used in this research applies to this rule, but profile curvature is reversed; negative profile curvature indicates convex areas and positive profile curvature indicates concave areas (ESRI, 2001; Moore *et al.*, 1991; Peschier, 1995).

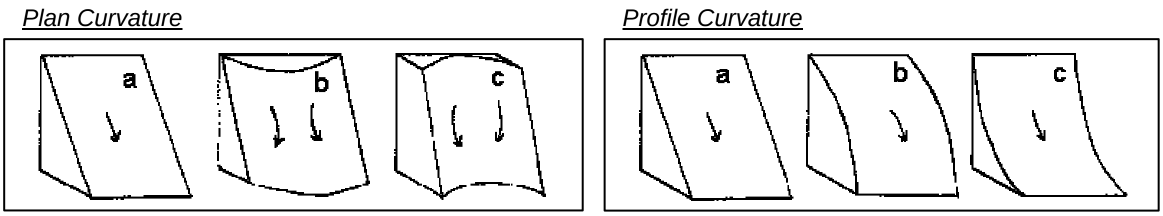


Figure 8: Possible curvatures; from left to right: plane (a), convex (b) and concave (c) curvature. The arrows indicate flow; convex plan curvature leads to diverging flow, concave plan curvature results in converging flow. Profile curvature affects the deceleration (b) and acceleration (c) of flow (Peschier, 1995).

3.2.2 *Predictive model performance*

The coefficients (B) of the full model can be used to state the full model as a mathematical equation in which the probability on the occurrence of forest is the function of the significant topographic and environmental variables (Hosmer & Lemeshow, 1989):

$$\text{P(forest)} = \frac{e^{(42.69393+(-0.01311\cdot \text{Height})+(-0.50856\cdot \text{CTI})+(-0.75409\cdot \text{AspectEW})+(0.06243\cdot \text{Slope})+\ldots}}{1+e^{(42.69393+(-0.01311\cdot \text{Height})+(-0.50856\cdot \text{CTI})+(-0.75409\cdot \text{AspectEW})+(0.06243\cdot \text{Slope})+\ldots}}$$

$$\frac{\ldots+(0.00701\cdot \text{STCI})+(-0.24278\cdot \text{AspectNS})+(3.19912\cdot \text{PRR})+(-0.95960\cdot \text{PlanCurv})}{\ldots+(0.00701\cdot \text{STCI})+(-0.24278\cdot \text{AspectNS})+(3.19912\cdot \text{PRR})+(-0.95960\cdot \text{PlanCurv})}$$

Table 2 presents an overview of the performance of the model in predicting the ‘*development area*’ per step of the logistic regression analysis. The 2 x 2 cross-tabulation table compares the observed response classes (forest/no forest) with the predicted response classes. The underlined numbers on the diagonal show the correct predictions (SPSS, 2003; Wuensch, 2005). The percentage in the row in which the observed forest distribution is *no-forest*, represents the *specificity* of the model. The row, in which the observed forest distribution is *forest*, represents the *sensitivity* of the model (Garson, 1998). Note that the percentages correctly classified pixels in these rows are approximately the same in all steps, with the sensitivity somewhat smaller than the specificity. This means that the logistic model has homoscedasticity (i.e. the variance of the variable forest/no-forest is the same for all data) and the model slightly under-predicts the occurrence of forest (Garson, 1998).

Table 2: Predictive capabilities of logistic model
This table shows the predictive accuracy of the logistic model per step in the stepwise logistic regression analysis. The underlined numbers show the correct predictions. The percentage in the row with no-forest observed represents the specificity of the model. The percentage in the row with forest observed represents the sensitivity of the model. Step 8 shows the predictive capabilities of the full model.

	Observed	Predicted		
Step	**Forest distribution**	**Forest distribution**		**Percentage Correct**
		No-Forest	Forest	
Step 1 (**Height**)	No-Forest	*52915*	11461	82.2
	Forest	12606	*46795*	78.8
		Overall Percentage		**80.6**
Step 2 (**CTI**)	No-Forest	*54589*	9787	84.8
	Forest	11491	*47910*	80.7
		Overall Percentage		**82.8**
Step 3 (**AspectEW**)	No-Forest	*54941*	9435	85.3
	Forest	11139	*48262*	81.2
		Overall Percentage		**83.4**
Step 4 (**Slope angle**)	No-Forest	*54891*	9485	85.3
	Forest	10668	*48733*	82.0
		Overall Percentage		**83.7**
Step 5 (**STCI**)	No-Forest	*55007*	9369	85.4
	Forest	10584	*48817*	82.2
		Overall Percentage		**83.9**
Step 6 (**AspectNS**)	No-Forest	*55032*	9344	85.5
	Forest	10499	*48902*	82.3
		Overall Percentage		**84.0**
Step 7 (**PRR**)	No-Forest	*55030*	9346	85.5
	Forest	10448	*48953*	82.4
		Overall Percentage		**84.0**
Step 8 (**Plan curvature**)	No-Forest	*55019*	9357	85.5
	Forest	10429	*48972*	82.4
		Overall Percentage		**84.0**

We can see a clear increase in the predictive capabilities of the model in the first steps. However, after step 6 the overall predictive capability of the model increases by less than 0.05 %. The effect of PRR and Plan curvature on the occurrence of forest, even though significant according to the -2LL-method and Wald statistic, can therefore be disputed. At least, for predictive importance these variables have almost no value.

Table 3 below shows a summary of the prediction accuracy in the *'development area'*, *'cross-validation area'* and *'test area'*. In the *'development area'* and *'cross-validation area'* the specificity of the model was somewhat higher than the sensitivity; in the *'test area'*, the specificity of the model is clearly much lower than the sensitivity. So the outcome of the model in the *'test area'* gives a large over-prediction of forest.

Table 3: Summary of predictive accuracy
This table shows a summary of the predictive accuracy of the full model in the different areas. The predictive accuracy in the 'cross-validation area' is comparable with the 'development area'. The test area shows a large over-prediction of forest; the sensitivity is much higher than the specificity (see table 2).

	Observed	Predicted		
Area	**Forest distribution**	**Forest distribution**		**Percentage Correct**
		No-Forest	Forest	
Development area	No-Forest	*55019*	9357	85.5
	Forest	10429	*48972*	82.4
		Overall Percentage		**84.0**
Cross-validation area	No-Forest	*55970*	11056	83.5
	Forest	9704	*47602*	83.1
		Overall Percentage		**83.3**
Test area	No-Forest	*59562*	40375	59.6
	Forest	4290	*75655*	88.8
		Overall Percentage		**74.2**

Figure 9a visualizes the predictions of the model in its own area ('*development area*') as well as in the '*cross-validation area*'. Figure 8b compares the prediction with the real forest distribution (the classification). Green colours indicate areas where forest is predicted where in fact forest was *not* present. Yellow colours indicate areas where no-forest is predicted where in fact forest *was* present.

In figure 9a, the areas indicated by the blue and green circle clearly show the importance of the inclusion of CTI into the model. Although the model correctly predicts the absence of forest on the wet valley bottom, the negative effect of CTI on the occurrence of forest is in fact even stronger than the model predicts; the green colours in figure 9b indicate the larger no-forest area on these valley bottoms. The area inside the green circle also shows the effect of the variable Aspect (east-west) on the prediction of forest. The left strip of forest (eastern aspect) is clearly smaller compared to the right strip of forest (western aspect). The upper limit of the right strip is located about 20 to 30 meters higher.

Figure 10a shows the model outcome in the '*test area*'. Figure 10b shows the predictive performance of the model. Very striking is the over-prediction of forest, especially in the southern, northern and western areas indicated by the green circles. This could well be the result of human impact on the area.

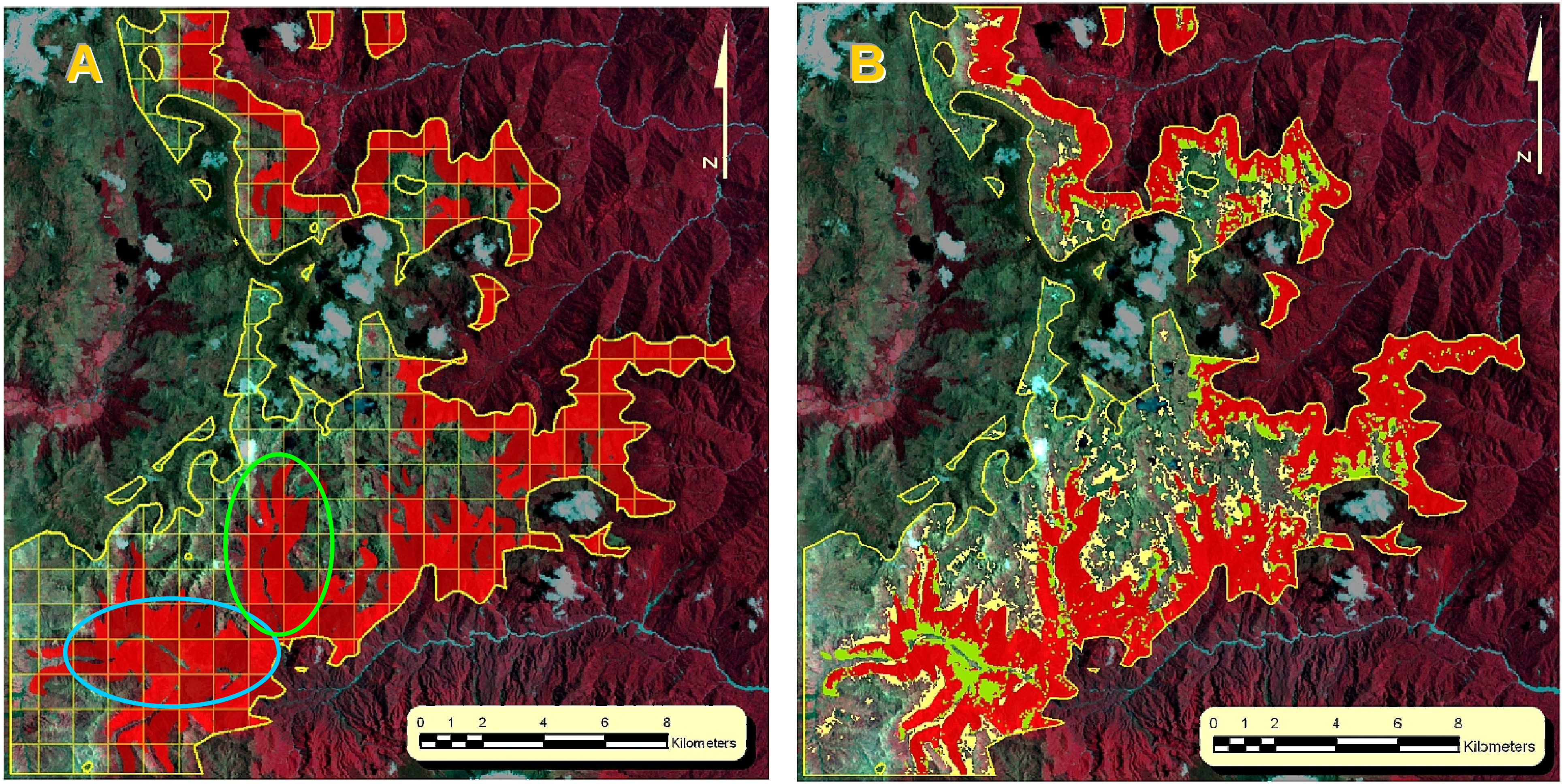


Figure 9: *Logistic model in 'development area' and 'cross-validation area'.*

A) The predictions of the model in its own area ('development area'; light blocks) as well as in the 'cross-validation area' (dark blocks). Red colours indicate forest. The highlighted area surrounded by the yellow border indicates the treeline zone.

B) Comparison of the prediction of the model and the real forest distribution (the classification). Green colours indicate areas where forest is predicted where in fact forest was not present. Yellow colours indicate areas where no-forest is predicted where in fact forest was present. Red colours indicate correctly predicted forest.

Overall classification accuracy in 'development area': 84.0 %; in 'cross-validation area': 83.3 %

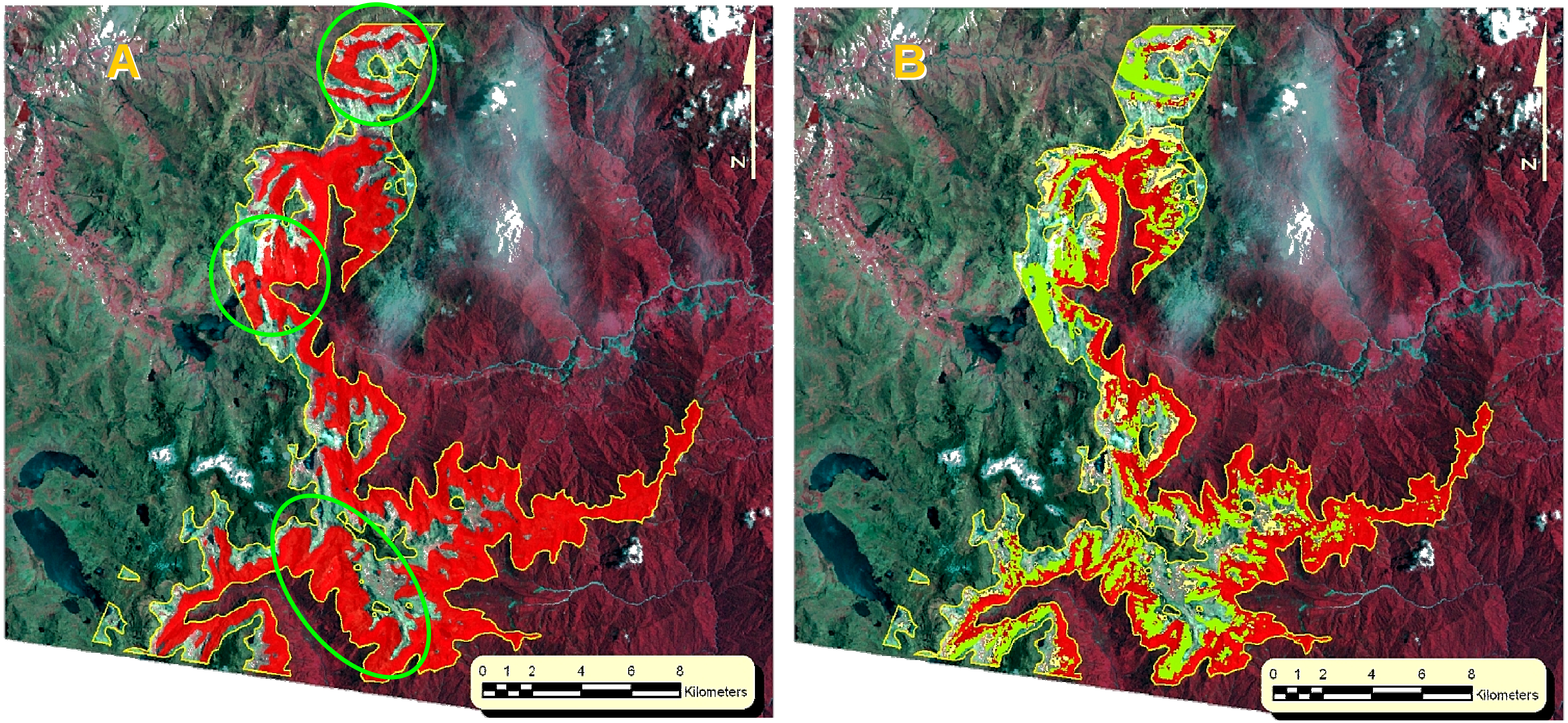


Figure 10: *Logistic model in 'test area'.*

A) The predictions of the model in the test area. Red colours indicate forest. The highlighted area surrounded by the yellow border indicates the treeline zone.

B) Comparison of the prediction of the model and the real forest distribution (the classification). Green colours indicate areas where forest is predicted where in fact forest was not present. Yellow colours indicate areas where no-forest is predicted where in fact forest was present. Red colours indicate correctly predicted forest.

Overall classification accuracy: 74.2%

3.2.3 *Importance of variables in 'full model'*

Exp(B) expresses the potential effect that the variables have on forest distribution. However, the importance of these effects for the model results differs. Therefore, we had a look at the measures that quantify this importance.

Table 4 shows a summary of the change in -2LL for every variable if it would be removed from the full model. The variables are ordered from high to low change in -2LL. The change in –2LL decreases as the variables decrease in importance. This statistic clearly shows the high importance of the variables height, aspect (east-west), and CTI (see table 4 in bold).

According to this method, even the low change of 82.78 of plan curvature is significant i.e. adding plan curvature to the model significantly increases the ability of the model to predict the occurrence of forest (Wuensch, 2005).

Table 4: Change in -2 Log Likelihood

This table gives an indication of the relative importance of the variables in the 'full model'. The variables are ordered from high to low importance based on the change in -2LL. The variables height, aspectEW and CTI show a clear impact on -2LL and are therefore displayed in bold.

Step	Variable	Change in -2 Log Likelihood	Change in percentages	df	Sig. of Change
Step 8	**Height**	**75950.79**	**90.58**	1	0.00
Full model	**AspectEW**	**3401.21**	**4.06**	1	0.00
	CTI	**3001.21**	**3.58**	1	0.00
	Slope	498.68	0.59	1	0.00
	AspectNS	345.18	0.41	1	0.00
	STCI	328.65	0.39	1	0.00
	PRR	237.19	0.28	1	0.00
	Plan curv	82.78	0.10	1	0.00

Another statistic used as an indication of the importance of the variables, is the Wald statistic of the full model. The Wald statistic can be found in table 1 on page 3. The outcome is similar to the outcome of the -2LL statistic.

3.2.4 *Stripped models*

Table 5 below shows a matrix with the bivariate correlation values of the explanatory variables in the full model. The highest correlated value is underlined. When taking a strict correlation criterion of 0.90, none of the variables are of concern when considering multicollinearity. However, the high correlation of -0.89 between Slope and PRR shows that we have to be cautious; much of the unexplained variance which could be explained by PRR is already explained by including slope angle into the model.

Table 5: Bivariate correlation of explanatory variables
This matrix shows the bivariate correlation of the explanatory variables in the 'full model'. Positive or negative values close to 1 indicate high correlation and were used to detect signs of bivariate multicollinearity. The highest correlation value (PRR vs. Slope) is underlined.

Variables	Height	Aspect EW	CTI	Slope	STCI	Aspect NS	PRR	Plan Curv
Height	1	-0.06	-0.24	0.10	-0.07	-0.06	-0.08	0.06
AspectEW	**-0.06**	1	-0.09	0.20	0.14	-0.04	-0.28	0.01
CTI	**-0.24**	**-0.09**	1	-0.57	0.08	-0.06	0.46	-0.57
Slope	**0.10**	**0.20**	**-0.57**	1	0.53	0.10	-0.89	0.07
STCI	**-0.07**	**0.14**	**0.08**	**0.53**	1	0.07	-0.48	-0.36
AspectNS	**-0.06**	**-0.04**	**-0.06**	**0.10**	**0.07**	1	0.09	0.00
PRR	**-0.08**	**-0.28**	**0.46**	**<u>-0.89</u>**	**-0.48**	**0.09**	1	-0.05
Plan Curv	**0.06**	**0.01**	**-0.57**	**0.07**	**-0.36**	**0.00**	**-0.05**	1

Next to the signs of bivariate multicollinearity we also had a look at the signs of multivariate multicollinearity. Table 6 below shows the multicollinearity diagnostic statistics. For these diagnostics we applied a VIF (Variance Inflation Factor) threshold value of 2.5 for indicating multicollinearity. In this statistic we see a very high inflation factor for slope angle and also PRR and CTI are well above the 2.5 criterion. For STCI we have to be cautious since its value is very close to the value of 2.5.

Table 6: Multicollinearity diagnostics
This table shows the multicollinearity diagnostics of the explanatory variables in the 'full model'. Tolerance shows the proportion of variance not explained by the other variables; high values are favorable. VIF (Variance Inflation Factor; = 1/Tolerance) is a measure of the times the variance of the variable coefficient is increased due to multicollinearity; low values are favorable. VIF was used to detect signs of multivariate multicollinearity. Explanatory variables in bold show signs of multicollinearity.

Variable	Tolerance	VIF
Slope	0.10	**10.03**
PRR	0.17	**6.03**
CTI	0.22	**4.48**
STCI	0.41	**2.46**
Plan Curv	0.54	1.85
Prof Curv	0.56	1.78
AspectNS	0.83	1.20
AspectEW	0.90	1.11
Height	0.91	1.10

Based on these assessments we decided to create stripped down models containing all the independent variables, but only one of the four variables showing signs of multivariate multicollinearity (variables in bold). The stripped models, with decreased multicollinearity, allow a better insight in the importance and effects of the individual variables. A summary of the full model and the stripped models is found in table 7 on the next page. The models are ordered from high to low goodness of fit. The variables are ordered from high to low change in -2LL.

The chi-squares and predictive accuracies of all stripped models are lower than these of the 'full model'. However, the stripped models allow us a better insight in the importance of the independent variables; the stripped models show no signs of bivariate or multivariate multicollinearity.

The stripped 'slope-model' represents a basic topographic model without any of the more complex environmental indices included. The importance of slope is suddenly much higher compared to its importance in the full model. The importance of PRR and STCI also increases dramatically in the stripped models. CTI, which already had quite an important role in the full model, also increases in importance and becomes the second most important variable, leaving AspectEW behind.

If we look at the Exp(B) of the variables in the stripped models we can see that CTI, slope and STCI have the same effect on the occurrence of forest as in the full model, only with a slight increase in effect. However, the effect of PRR in the stripped model is reversed: high radiation now causes a decrease in forest probability.

In the full model and in the stripped 'CTI-model', concave plan and profile curvatures cause a higher probability of forest occurrence. In the remaining stripped models the effect is reversed: concave plan and profile curvatures cause lower probability of forest occurrence.

Table 7: Stripped models

This table shows a summary of the 'full model' and of the stripped models which contain all variables but only one of the four variables showing signs of multivariate multicollinearity (variables in bold). The models are ordered from high to low goodness-of-fit (Chi-Square). The variables are ordered from high to low change in -2LL. Exp(B) shows the coefficient related to the variable. Change in -2LL (and change in percentages) indicates the relative importance of the variable in the model.

Model	Variable	Exp(B)	Change in -2 Log Likelihood	Change in percentages	df	Sig. of the Change
Full Model Total Chi-Square: **79698.95** Overall accuracy: **84.0**	Height	0.98697	75950.79	90.58	1	0.00
	AspectEW	0.47044	3401.21	4.06	1	0.00
	CTI	0.60136	3001.21	3.58	1	0.00
	Slope angle	1.06442	498.68	0.59	1	0.00
	AspectNS	0.78444	345.18	0.41	1	0.00
	STCI	1.00703	328.65	0.39	1	0.00
	PRR	24.51099	237.19	0.28	1	0.00
	Plan	0.38305	82.78	0.10	1	0.00
CTI Total Chi-Square: **77526.64**	Height	0.98714	76238.58	83.83	1	0.00
	CTI	**0.50335**	**10522.35**	**11.57**	1	0.00
	AspectEW	0.50105	3207.61	3.53	1	0.00

Overall accuracy: **83.6**	Plan	0.06106	837.78	0.92	1	0.00
	AspectNS	0.88631	101.35	0.11	1	0.00
	Prof	1.66804	35.72	0.04	1	2.28 E-09
Slope Total Chi-Square: **76682.92** Overall accuracy: **83.5**	Height	0.98766	73975.32	83.06	1	0.00
	Slope	**1.10086**	**9678.62**	**10.87**	1	0.00
	AspectEW	0.44146	4343.49	4.88	1	0.00
	Prof	0.11988	641.31	0.72	1	0.00
	Plan	3.44700	212.10	0.24	1	0.00
	AspectNS	0.84152	207.30	0.23	1	0.00
PRR Total Chi-Square: **72722.11** Overall accuracy: **82.7**	Height	0.98840	70759.06	86.68	1	0.00
	PRR	**0.00118**	**5717.82**	**7.00**	1	0.00
	AspectEW	0.44661	4254.13	5.21	1	0.00
	Prof	0.15388	528.35	0.65	1	0.00
	Plan	4.89414	370.46	0.45	1	0.00
	AspectNS	1.03008	6.42	0.01	1	0.01
STCI Total Chi-Square: **70235.83** Overall accuracy: **82.2**	Height	0.98919	66404.95	88.88	1	0.00
	STCI	**1.01715**	**3231.53**	**4.33**	1	0.00
	AspectEW	0.53423	2830.62	3.79	1	0.00
	Plan	51.25384	1985.04	2.66	1	0.00
	Prof	0.32185	196.19	0.26	1	0.00
	AspectNS	0.90961	66.88	0.09	1	3.33 E-16

The above illustrates the impact of the wet, non-forest areas (explained by CTI) on the effect (Exp(B)) of the other independent variables predicting forest/non-forest. When the variable CTI is not included in the model, other variables try to take over its negative effect on forest probability:

- The effect and importance of slope angle (higher slope angle = higher forest probability) increases since the non-forest areas related to wetness are mainly located on the lower slope angles.
- The same goes for STCI (higher STCI = higher forest probability).
- The effect of PRR is reversed (higher PRR = lower forest probability) and its importance increases since the lighter areas occur mainly in the valley bottoms, in gullies and on ridges.
- The effect of plan and profile curvature is reversed (more concavity = lower forest probability) since the non-forest areas related to wetness are mainly located in concave areas. Slope and PRR have apparently some overlap with plan curvature; the importance of plan curvature in these models is not very high. The importance of plan curvature in the 'STCI-model' increases considerably since STCI is not capable of explaining all of the non-forest areas related to wetness: STCI is generally high on steep slopes (low CTI, high forest cover), but can also be high in wet

gullies with low forest cover – this low forest cover then needs to be accounted for by plan curvature.

3.2.5 *Importance of variables for forest distribution at alpine treeline*

When taking the above theory into account we can make a summary of the importance of the variables in explaining forest distribution in the alpine treeline zone. Table 8 below shows the variables from high to low importance based on the full model and based on the stripped models. Plan and profile curvature are not included in this list because of there generally low model importance.

Table 8: Importance of explanatory variables
This table shows a summary of the importance of the variables in explaining forest distribution at alpine treeline. The old arrangement is based on the importance of the variables in the 'full model'. The new arrangement is an assessment of the importance of the variables after analyzing the stripped models.

	Old arrangement (based on full model)		**New arrangement** (based on stripped models)
1.	Height	1.	Height
2.	AspectEW	2.	CTI
3.	CTI	3.	AspectEW
4.	Slope	4.	STCI
5.	AspectNS		
6.	STCI		
7.	PRR		

The new arrangement is based on the information of the stripped models and taking environmental effects into account. First of all, we can say that CTI is better in explaining the non-forest areas related to wetness compared to Slope, PRR and STCI. Therefore we can regard AspectEW as the third, most important variable because it is explaining a different environmental effect. This corresponds to the outcome of visual analysis of forest distribution at different altitudes. At the lower altitudes (below average) we can see no clear preference of forest for a certain aspect. But at higher altitudes (above average) we can see a clear preference of forest for western slopes (see figure 11).

Secondly, slope is highly related to CTI. The importance of CTI decreases dramatically when introducing slope into the model. A model with all the variables included except slope angle has only a slightly smaller goodness of fit (see table 9). Its overall predictive accuracy remains at 84%. In other words, slope tries to explain much of the areas already explained by CTI, thereby not introducing any new environmental information. Its importance to the model is very limited. Therefore slope angle is not included in the new arrangement.

Thirdly, in the model with only slope excluded, STCI is a quite important variable responsible for almost 2 % change in -2LL without greatly decreasing the importance of CTI; i.e. STCI introduces important new environmental information.

Finally, AspectNS has almost no change in -2LL in this model and PRR is not significant enough to be included. Therefore, these variables are not included in the new arrangement.

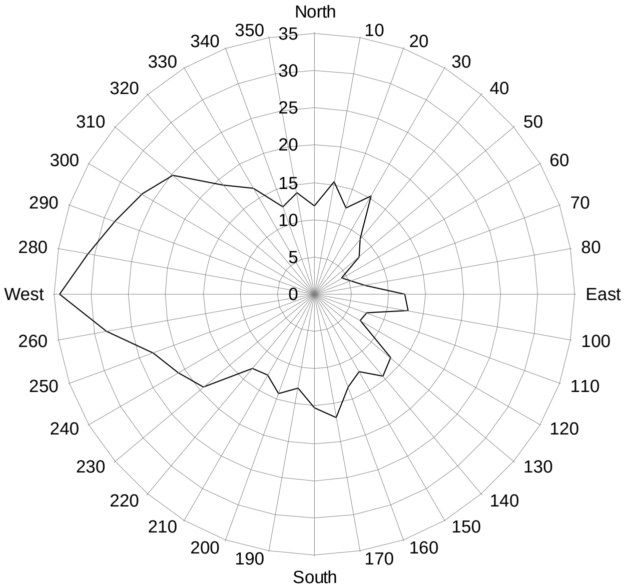


Figure 11: Aspect ecogram

The ecogram shows the relative forest distribution in the 'development area' above the average treeline altitude (3634 meters) along a circular aspect gradient (in degrees). The radial axis represents the amount of forest (in %) found on a certain aspect. The graph clearly illustrates that at high altitudes, forest is mainly located on western slopes.

Table 9: Stripped model (slope exclusion)
This table shows a summary of the stripped model containing all the variables except slope. The variables are ordered from high to low change in -2LL. Exp(B) shows the coefficient related to the variable. Change in -2LL (and change in percentages) indicates the relative importance of the variable in the model. Variables in bold are taken up in the new arrangement of variable importance.

Model	Variable	Exp(B)	Change in -2 Log Likelihood	Change in percentages	df	Sig. of the Change
Model with Slope excluded Total Chi-Square: **79297.76** Overall accuracy: **84.0**	**Height**	-0.01302	75313.28	**83.37**	1	0
	CTI	-0.65702	9061.93	**10.03**	1	0
	AspectEW	-0.77011	3827.53	**4.24**	1	0
	STCI	0.01227	1771.11	**1.96**	1	0
	AspectNS	-0.16004	173.44	0.19	1	0
	Prof	0.89710	104.27	0.12	1	0
	Plan	-1.01340	88.57	0.10	1	0

4 Discussion

This chapter deals with interpreting the outcome of the logistic model, investigating model stability and understanding the limitations and advantages of using a logistic modeling approach.

4.1 Model interpretation

Logistic regression analysis is a useful tool to assess the relative importance of variables for vegetation distribution as long as the effects of multicollinearity are taken into account. Despite the possibility to discriminate the importance of the variables under research, the underlying ecological mechanisms are still hard to quantify. The four identified variables having a pronounced effect on forest distribution at alpine treeline (height, CTI, aspectEW and STCI) may influence several edaphic, physical and climatic factors. We also have to keep in mind that we deal with statistical relationships which are not necessarily good replacements for actual ecological manifestations (Hoersch *et al.*, 2002).

The results of this study clearly showed the importance of altitude (variable height) as influencing factor, responsible for approximately 85% of explained variance. Altitude is an indicator of several ecological gradients which may influence forest distribution at alpine treeline; e.g. temperature and precipitation (Bian & Walsh, 1993; Dirnböck *et al.*, 2003). In the study area, temperature gradient, influenced by altitude, will probably be the most important determinant for the spatial distribution of forest.

CTI turned out to be another important variable influencing forest distribution at alpine treeline, responsible for approximately 11% to 12% of explained variance. The negative effect of CTI can be related to excessive soil moisture conditions and drainage areas with high erosive power. Another possibility is the effect of air, cooling down during the night at higher altitudes, sliding down the slopes and causing a blanket of stagnant cool air in the narrow valleys (Paulsen & Korner, 2001; Wardle, 1985). Areas with potential susceptibility to this effect coincide with areas having high CTI (indicating areas with excessive soil moisture).

AspectEW (east-west) was the third variable to show a pronounced effect on forest distribution at alpine treeline, responsible for approximately 4.25% of explained variance. The ecogram indicated that, at higher altitude, forest has a preference for western slopes, which suggests that forest on western slopes can reach higher elevations. Slope aspect is related to several biophysical factors, e.g. incoming solar radiation and humidity (Dirnböck *et al.*, 2003; Körner, 1998). During the morning, when the sun is located in the east, the study area is often practically cloud free. In the afternoon, clouds are rising up from the Amazon basin into the study area, limiting the incoming solar radiation. The possibility exists that this cloud-provided shelter serves as an ecological niche, allowing tree establishment beyond the shelter of trees, and thereby positively influencing the advance of western treeline to higher altitude. In contrast, harsh sunlight on eastern slopes may limit tree establishment at open spaces beyond the treeline, preventing the advance of the eastern treeline to higher altitudes (Smith & Young, 1987).

STCI was the last variable to show a pronounced effect on forest distribution at alpine treeline, responsible for approximately 2% of explained variance. The actual biophysical effect is hard to assess. The variable shows a positive effect on alpine forest distribution, but it is rather awkward to conclude that forest at alpine treeline prefers areas of high potential erosion. It is more likely that STCI has as small correlation with a biophysical factor (e.g. soil condition) that positively influences forest establishment.

The remaining factors, slope, aspectNS (north-south), profile and plan curvature did not show a distinct effect on forest distribution at the current scale. This does not mean that these variables have no influence at all. It only implicates that in our model, at the current scale of investigation, the other variables are better in explaining forest distribution at alpine treeline.

Other studies attempting to explain vegetation distribution at comparable scales by means of topographical variables, experienced similar difficulties in discriminating causality effects. Comparison with most of this research is rather limited since these studies focused mainly on mid- and high latitude alpine regions (Allen & Walsh, 1996; Baker *et al.*, 1995; Horsch, 2003; Virtanen *et al.*, 2004; Walsh *et al.*, 2003). Allen & Walsh (1996), for example, investigated spatial pattern of alpine treeline in Glacier National Park, Montana. They found strong significance with elevation and related the effect to a temperature gradient. Other significant factors included slope angle and solar radiation potential. These were related to snow avalanches, debris flows and the location of permanent snowfields (Allen & Walsh, 1996). In our equatorial study region the effects of snow avalanches and permanent snowfields on forest distribution are of no importance, which may explain why we did not find similar relationships.

Brown (1994) constructed linear regression models for Glacier National Park, Montana, to predict vegetation types by using topographical and biophysical disturbance variables. He also found elevation to be the most predictive factor, followed by solar radiation, snow accumulation and topographic soil moisture potential. Baker *et al.* (1995), who performed their research in Rocky Mountain National Park, Colorado, and Virtanen *et al.* (2004), who investigated the artic treeline in Northeast Russia, stated the negative effect of topographic soil moisture on vegetation distribution at treelines. Baker *et al.* (1995) found the effect of wetland interruptions especially important in the lower regions, while the effect of lake interruptions increased at higher elevations. Virtanen *et al.*

(2004) found the effect to be significant at a spatial scale of 3 kilometers and increasing in significance at higher resolutions. This underlines the importance of incorporating topographic wetness in alpine vegetation modeling.
Recent research by Hoch & Korner (2005), who studied *Polylepis* species at alpine treeline in the Oruro province, Bolivia, found aspect related limitations of the species at southern slopes. Above 4400 meters altitude, *Polylepis* was mainly located at the warmer and drier northern slopes, facing the equator. In our study area, located approximately 2º south of the equator, a north-south influence was not found which can be explained by the fact that their study area is located at 18º south of the equator.
Finally, Hörsch (2003) investigated 25 DEM derived variables to model the spatial distribution of forest in the central Alps. She found all topographical variables (elevation, slope, aspect, radiation, curvature, etc.) to be significant in explaining vegetation alliances. However, she did not take into account the effects of spatial autocorrelation. At the current scale of analysis we expected spatial similarity to be caused mainly by similar topographical position and therefore did not include any terms in our model which accounted for spatial autocorrelation. However, when using DEM-derived variables, data points are located relatively close to each other, thus violating the assumption of independent variables. In this case, pseudo-replication occurs, affecting the actual number of degrees of freedom (Heffner *et al.*, 1996). This causes model coefficients to increase in statistical significance which is probably the reason why Hörsch found all topographical variables to be significant in explaining alpine vegetation alliances. When choosing not to incorporate spatial autocorrelation it is essential to treat the significance of model coefficients carefully. While modeling alpine forest distribution at higher spatial resolutions (e.g. scale of individual trees) it is recommended to incorporate spatial autocorrelation into the analysis, thereby taking effects such as local dispersion and sheltering into account; e.g. by means of iterative methods (Augustin *et al.*, 1996) or by taking a subsample of data points thereby increasing distance and decreasing correlation between points (Brenning, 2005; Brown, 1994).

Logistic regression analysis lacks the ability to make firm statements about the actual percent of variance explained by the topographical and topography derived variables (Garson, 1998), but the predictive accuracies suggest that topographical and topography derived variables are capable of explaining a significant amount of alpine forest distribution. This is promising for the use of topographic variables in ecological modeling as a substitute for biophysical factors directly influencing vegetation distribution, especially in areas with large topographic variation. Moreover, previous research in regions with high heterogeneous geomorphology suggested that topographical factors were more useful in predicting vegetation communities than spectral attributes (Dirnböck *et* *al.*, 2003). Furthermore, compared to direct influencing factors, topography and topography related factors are relatively inexpensive and easy to obtain (Dirnböck *et al.*, 2003; Gottfried *et al.*, 1998).

However, topography and topography derived variables are not capable of accounting for all biophysical factors influencing forest distribution at alpine treeline. The variable PRR for example is an index identifying potential relative radiation. The calculation assumes clear sky and no atmospheric scattering effects, which is often not the case, thereby limiting the value of this factor (Pierce *et al.*, 2005). Combined with the location of the study area near the equator may be the reason why we did not find this variable to have a pronounced effect on alpine forest distribution. Also the effect of human landuse has to be taken into account. Human landuse can overrule the effects of topography (Dirnböck *et al.*, 2003).

4.2 Model applicability and model stability

The full logistic model had an overall predictive accuracy of 83.3 % in the '*cross-validation area*' and 74.2 % in the '*test area*'. It is clear that, at the current scale of investigation, topography has a significant influence on forest distribution at alpine treeline. This influence is mainly limited to constraints determined by altitude, CTI, aspect and STCI. There is a significant difference in scale between the data of forest distribution and the data of topographic variables It is important to note that resampling the DEM from 90 to 28.5 meters results in a certain degree of pseudo-resolution. Combined with the smoothing caused by artifact removal, the DEM is not capable of indicating micro-scale variation and small depression areas which could be of importance to determine forest distribution at alpine treeline (Dirnböck *et al.*, 2003; Horsch, 2003). Other biophysical factors not explained by topography may also have a significant influence. The current scale of analysis is probably the limit to use topographic and topography derived variables to predict forest distribution at alpine treeline. At larger scales the predictive power of topography decreases significantly (Guisan & Zimmermann, 2000). A DEM with a spatial resolution comparable to that of the Landsat image would probably have increased predictive accuracy.

Model stability is indicated by several factors. Quality and accuracy of the input data is a first important factor which has to be taken into account. After orth-rectification, the Landsat ETM+ image has absolute horizontal accuracy of +/- 64 meters (GLCF, 2006). The SRTM DEM has horizontal absolute and relative accuracy of +/- 20 respectively 15 meters. The vertical absolute and relative accuracy is +/- 16 respectively 6 meters (Rodriguez *et al.*, 2005; USGS, 2003). The DEM contained several gaps with missing values. In SRTM data these gaps can be found especially along rivers, in lakes and in areas with high slope angle (particularly in the Himalaya and the Andes) (Jarvis *et al.*, 2004). Filling the gaps by interpolating neighboring values creates pseudo elevational data. The actual elevations in these regions are unknown which may be a cause of error. (Falorni *et al.*, 2005) found vertical accuracy of SRTM data influenced by high-relief terrain. In these areas, elevation errors and the amount of missing values increased significantly. It has also been found that SRTM data suffers from a systematic error related to aspect. Elevation of (north-)eastern slopes is overestimated compared to elevation of western slopes. This phenomenon has been directly related to the incidence angle in the original radar images caused by the flight path direction of the space shuttle (Jarvis *et al.*, 2004).

Local displacements influenced by the errors as stated above may cause error in overlap between the Landsat-derived forest distribution and the topographic variables derived from the SRTM DEM. However, the variables in our model have shown to mainly exert their influence on a broad spatial scale, possibly due to the lack of micro-scale variation in the DEM. Minor displacement of the overlapping variables can therefore be accepted. Moreover, the relationship between AspectEW (east-west) and forest distribution can not be explained by the aspect-related DEM error. The outcome of our model result indicates that forest reaches higher altitudes at western slopes. If the aspect-error had significant impact, this relationship would be reversed; i.e. forest would have a preference for eastern slopes.

Finally, it must be noted that despite the fact that forest in the Landsat image was well recognizable and the classification result showed large similarities with forest in the original image, there will be some classification error involved, influencing predictive accuracy.

Other factors determining model stability are the statistical method used to build the model and the applicability of the model in other areas.

Compared to other statistical methods (e.g. discriminant analysis) logistic regression has been found to be a relatively robust method (Garson, 1998). However, this approach assumes equilibrium between alpine forest distribution and biophysical factors influenced by topography. The model is therefore not suitable for application in rapidly changing environments. On the other hand, the model has been applied in an alpine environment in which vegetation has been found to react slowly to changing environmental conditions (Guisan & Zimmermann, 2000). Therefore, static logistic models are well-suited for alpine environments. But it is important to note that the application of the model is limited to a rather small geographic extent because of differences in direct influencing factors (e.g. climate and resource gradients). Plant species are forced to adapt to these differences by selecting topographic positions with favorable conditions. In other regions, the same topographic positions can cause different effects on the direct influencing factors (Guisan & Zimmermann, 2000). Furthermore, since one of the possible applications of this method is to indicate human impact, this method needs calibration areas where the effects of human landuse are negligible.

The '*cross-validation area*' allowed a first investigation of the stability of the logistic model in a similar area. The '*development area*' and '*cross-validation area*' are practically not spatially dependent since a checkerboard pattern with blocks of +/- 1 x 1 km was used to hold data back. This supports good comparison; the '*development area*' has almost no spatial effect on the '*cross-validation area*'. The predictive accuracies in the '*cross-validation area*' (84.0%) compared to the '*development area*' (83.3%) suggest that the model is quite robust. The lower accuracy of 74.2% in the '*test area*' is probably caused by higher disturbance caused by human influence resulting in an over-prediction of forest. For external validation, this area is therefore not very suitable, but it clearly shows the influence of changed conditions (in this case: human disturbance) on the spatial distribution of forest. For further investigation of model stability it is advised to locate an additional area similar to the '*development area*' to externally validate the model. Constructing the model with varying extents of the treeline zone (e.g. 1.0, 1.5, and 2.5 standard deviations) may also contribute to further insight in model stability since varying extents can influence model performance and the relative importance of the variables (Fagan *et al.*, 2003; Virtanen *et al.*, 2004).

Finally, the selection of the variables resulting from the stepwise logistic regression procedure varied a lot. However, the variables height, aspectEW and CTI were always included in the model, showing a pronounced effect on forest distribution. These variables can therefore be regarded as robust variables in explaining forest distribution in our study area. Variable STCI is regarded as somewhat less robust since it showed a little more variation in importance, but this was mainly caused by the effects of multicollinearity. When these effects had been removed, STCI also showed a pronounced effect on alpine forest distribution.

5 Conclusions & Recommendations

The results of this study demonstrated the possibilities of using vegetation modeling to understand the factors influencing forest distribution at tropical alpine treelines. A Landsat ETM+ image was used to derive the spatial distribution of forest. An SRTM DEM was used to derive several topographical and topography-derived variables. These variables were used as a substitute for direct field measurements of direct influencing factors (e.g. climate and soil conditions). The data were used to construct a logistic regression model to predict forest distribution at alpine treeline and to investigate the effect and relative importance of these variables on the spatial distribution of forest.

The results of this study provide further insight in the factors influencing the spatial distribution of forest at alpine treelines. In our study area, topographical and topography-derived variables influencing forest distribution at landscape level are: altitude (height), topographic wetness (CTI), eastness (AspectEW) and erosion potential (STCI). This knowledge can be used for modeling efforts of tropical alpine vegetation in the future.

Quantification of the biophysical processes and conditions influenced by these variables was limited, due to the absence of supportive field measurements. However, the predictive accuracy of the model clearly indicated the high topographic impact on several biophysical factors influencing alpine forest distribution. Absence of a clear influence related to north-south aspect and potential radiation supported the theory of topography as a factor influencing a different set of biophysical factors in the tropics. Effects of topography in temperate regions influence for example snow accumulation and incoming radiation, but since these effects are absent in tropical regions, the additional effects of topography can be studied. Therefore, the results of this study underline the potential of tropical regions to provide further knowledge of the factors influencing alpine treelines.

The method developed during this research allows quick investigation of factors with potential influence on alpine forest distribution by using inexpensive and easy to obtain Landsat and DEM imagery. Furthermore, the method supports straightforward mapping of alpine forest distribution.

At the current scale of investigation, using only topography and topography-derived variables is not sufficient enough to provide a detailed prediction of the actual treeline position. However, the model is relatively easy to construct, compared to more advanced mechanistic models which require large numbers of parameters that are often not available. This model allows fast application in remote areas were many of these parameters are lacking. It has the possibility to quickly compare regional differences in forest distribution and give a first indication of degraded areas caused by either human impact or natural disasters (e.g. disease). More comprehensive methods (e.g. combination of environmental monitoring and field measurements) can be applied at the located areas to investigate the nature of this impact. Combined with automatic vegetation mapping, the method can also be valuable for planning and conservation issues (Dirnböck *et al.*, 2003).

To further develop the applicability and usefulness of the proposed method, we would like to make several recommendations.
First of all, next to the application of the model in the '*cross-validation area*' and the '*test area*', it is recommended to validate the model also in an additional area similar to the '*development area*'; i.e. an area without distinctive human influence. This would greatly improve the validation of the method, allowing firmer conclusions and greater model applicability (Guisan & Zimmermann, 2000).
Secondly, the method is based on one spatial scale and assumes equilibrium. To fully understand the landscape under study, a multi-scale, multi-temporal approach is recommended to fully capture the complicated topographic heterogeneity and temporal variability (Fagan *et al.*, 2003; Levin, 1992; Wu, 2004). Ecological modeling could benefit greatly from a multi-scale, multi-temporal approach to investigate the effects of topography and topography-derived variables on alpine vegetation. To really investigate causality effects the method should be complemented by supportive field measurements. Improved understanding of topographic causality effects can greatly improve predictive accuracy and applicability of mechanistic vegetation models. Progress in this understanding can be made especially in tropical regions, were the more renowned effects of temperate regions (e.g. north-south radiation difference) are absent. Improvements in remote sensing imagery, concerning both spatial and temporal resolution, can support multi-scale, multi-temporal logistic modeling approaches in remote environments in the near future.
Finally, investigating the potential of automatic classification techniques based on pattern analysis and boundary detection (Fagan *et al.*, 2003) as well as advanced visualization techniques, such as 3 dimensional mapping (Walsh *et al.*, 2003), would open up new possibilities of integrating these techniques into our method. This would significantly improve the ease of our method and simplify the evaluation of the model results making it more appropriate for planning purposes.

References

Allen, R.B., Hewitt, A.E., & Partridge, T.R. (1995) Predicting Land-Use Suitability from Vegetation and Landform in Depleted Semiarid Grassland, New-Zealand. *Landscape and Urban Planning*, **32**, 31-42.

Allen, T.R. & Walsh, S.J. (1996) Spatial and compositional pattern of alpine treeline, Glacier National Park, Montana. *Photogrammetric Engineering and Remote Sensing*, **62**, 1261-1268.

Armstrong, G.D. & Macey, A. (1979) Proposals for a Sangay National-Park in Ecuador. *Biological Conservation*, **16**, 43-61.

Augustin, N.H., Cummins, R.P., & French, D.D. (2001) Exploring spatial vegetation dynamics using logistic regression and a multinomial logit model. *Journal of Applied Ecology*, **38**, 991-1006.

Augustin, N.H., Mugglestone, M.A., & Buckland, S.T. (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, **33**, 339-347.

Baker, W.L., Honaker, J.J., & Weisberg, P.J. (1995) Using Aerial-Photography and Gis to Map the Forest-Tundra Ecotone in Rocky-Mountain National-Park, Colorado, for Global Change Research. *Photogrammetric Engineering and Remote Sensing*, **61**, 313-320.

Bian, L. & Walsh, S.J. (1993) Scale Dependencies of Vegetation and Topography in a Mountainous Environment of Montana. *Professional Geographer*, **45**, 1-11.

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

Brown, D.G. (1994) Predicting Vegetation at Treeline Using Topography and Biophysical Disturbance Variables. *Journal of Vegetation Science*, **5**, 641-656.

Butler, D.R., Hill, C., Malanson, G.P., & Cairns, D.M. (1994) Stability of alpine treeline in Glacier National Park, Montana, U.S.A. *Phytocoenologia*, **22**, 485-500.

Calef, M.P., McQuire, A.D., Epstein, H.E., Rupp, T.S., & Shugart, H.H. (2005) Analysis of vegetation distribution in Interior Alaska and sensitivity to climate change using a logistic regression approach. *Journal of Biogeography*, **32**, 863-878.

Camarero, J.J. & Gutiérrez, E. (2004) Pace and pattern of recent treeline dynamics: response of ecotones to climatic variability in the Spanish Pyrenees. *Climatic Change*, **63**, 181.

Daniels, L.D. & Veblen, T.T. (2004) Spatiotemporal influences of climate on altitudinal treeline in northern Patagonia. *Ecology*, **85**, 1284–1296.

De Roo, A.P.J. (1998) Modelling runoff and sediment transport in catchments using GIS. *Hydrological Processes*, **12**, 905-922.

delBarrio, G., Alvera, B., Puigdefabregas, J., & Diez, C. (1997) Response of high mountain landscape to topographic variables: Central Pyrenees. *Landscape Ecology*, **12**, 95-115.

Dirnböck, T., Dullinger, S., Gottfried, M., Ginzler, C., & Grabherr, G. (2003) Mapping alpine vegetation based on image analysis, topographic variables and anonical Correspondence Analysis. *Applied Vegetation Science*, **6**, 85–96.

Dirnböck, T., Hobbs, R.J., Lambeck, R.J., & Caccetta, P.A. (2002) Vegetation distribution in relation to topographically driven processes in south-western Australia. *Applied Vegetation Science*, **5**, 147-158.

Dullinger, S., Dirnböck, T., & Grabherr, G. (2004) Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. *Journal of Ecology*, **92**, 241-252(12).

ESRI (2001) ARC/INFO Help, Redlands, CA, USA.

Fagan, W.F., Fortin, M.J., & Soykan, C. (2003) Integrating edge detection and dynamic modeling in quantitative analyses of ecological boundaries. *Bioscience*, **53**, 730-738.

Falorni, G., Teles, V., Vivoni, E.R., Bras, R.L., & Amaratunga, K.S. (2005) Analysis and characterization of the vertical accuracy of digital elevation models from the Shuttle Radar Topography Mission. *Journal of Geophysical Research-Earth Surface*, **110**.

Felicisimo, A.M., Frances, E., Fernandez, J.M., Gonzalez-Diez, A., & Varas, J. (2002) Modeling the potential distribution of forests with a GIS. *Photogrammetric Engineering and Remote Sensing*, **68**, 455-461.

Freeman, T.G. (1991) Calculating Catchment-Area with Divergent Flow Based on a Regular Grid. *Computers & Geosciences*, **17**, 413-422.

Garson, D. (1998) Logistic regression. In http://www2.chass.ncsu.edu/garson/pa765/logistic.htm. North Carolina State University.

Gessler, P.E., Chadwick, O.A., Chamran, F., Althouse, L., & Holmes, K. (2000) Modeling soil-landscape and ecosystem properties using terrain attributes. *Soil Science Society of America Journal*, **64**, 2046-2056.

GLCF (2006) Global Land Cover Facility. In http://glcf.umiacs.umd.edu/index.shtml. Institute for Advanced Computer Studies, Maryland, USA.

Gottfried, M., Pauli, H., & Grabherr, G. (1998) Prediction of vegetation patterns at the limits of plant life: A new view of the alpine-nival ecotone. *Arctic and Alpine Research*, **30**, 207-221.

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

Hansen, A.J., Neilson, R.R., Dale, V.H., Flather, C.H., Iverson, L.R., Currie, D.J., Shafer, S., Cook, R., & Bartlein, P.J. (2001) Global change in forests: Responses of species, communities, and biomes. *Bioscience*, **51**, 765-779.

Heffner, R.A., Butler, M.J., & Reilly, C.K. (1996) Pseudoreplication revisited. *Ecology*, **77**, 2558-2562.

Hilbert, D.W. & Ostendorf, B. (2001) The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. *Ecological Modelling*, **146**, 311-327.

Hoch, G. & Korner, C. (2005) Growth, demography and carbon relations of Polylepis trees at the world's highest treeline. *Functional Ecology*, **19**, 941-951.

Hoersch, B., Braun, G., & Schmidt, U. (2002) Relation between landform and vegetation in alpine regions of Wallis, Switzerland. A multiscale remote sensing and GIS approach. *Computers, environment and Urban Systems*, **26**, 113-139.

Holtmeier, F.K. & Broll, G. (2005) Sensitivity and response of northern hemisphere altitudinal and polar treelines to environmental change at landscape and local scales. *Global Ecology and Biogeography*, **14**, 395-410.

Horsch, B. (2003) Modelling the spatial distribution of montane and subalpine forests in the central Alps using digital elevation models. *Ecological Modelling*, **168**, 267-282.

Hosmer, D.W. & Lemeshow, S. (1989) *Applied logistic regression* New York [etc.] : Wiley.

Hutchinson, M.F. (1989) A New Procedure for Gridding Elevation and Stream Line Data with Automatic Removal of Spurious Pits. *Journal of Hydrology*, **106**, 211-232.

Jarvis, A., Rubiano, J., Nelson, A., Farrow, A., & Mulligan, M. (2004). Practical use of SRTM data in the tropics – Comparisons with digital elevation models generated from cartographic data. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.

Jenness, J. (2005) Some Thoughts on Analyzing Topographic Habitat Characteristics. In http://fwie.fw.vt.edu/tws-gis/newsletter/jun_05/topo_char.pdf. Jenness Enterprises, Flagstaff, AZ, USA.

Kjällgren, L. & Kullmann, L. (1998) Spatial patterns and strcuture of the mountain birch tree-limit in the southern Swedish Scandes - a regional perspective. *Geogr. Ann.*, **80**, 1-16.

Körner, C. (1998) A re-assessment of high elevation treeline positions and their explanation. *Oecologia*, **115**, 445-459.

Kupfer, J.A. & Cairns, D.M. (1996) The suitability of montane ecotones as indicators of global climatic change. *Progress in Physical Geography*, **20**, 253-272.

Levin, S.A. (1992) The Problem of Pattern and Scale in Ecology. *Ecology*, **73**, 1943-1967.

Moen, J., Aune, K., Edenius, L., & Angerbjörn, A. (2004) Potential Effects of Climate Change on Treeline Position in the Swedish Mountains. *Ecology and Society*, **9**, 16.

Moore, I.D., Gessler, P.E., Nielsen, G.A., & Peterson, G.A. (1993) Soil Attribute Prediction Using Terrain Analysis. *Soil Science Society of America Journal*, **57**, 443-452.

Moore, I.D., Grayson, R.B., & Ladson, A.R. (1991) Digital Terrain Modelling: A Review Of Hydrological, Geomorphological, And Biological Applications. *Hydrol. Process.*, **5**, 3-30.

NREL (2001) Solar Position Calculator. In http://www.nrel.gov/. Midwest Research Institute, Batelle, USA.

Ohsawa, M. (1990) An Interpretation of Latitudinal Patterns of Forest Limits in South and East-Asian Mountains. *Journal of Ecology*, **78**, 326-339.

Paulsen, J. & Korner, C. (2001) GIS-analysis of tree-line elevation in the Swiss Alps suggests no exposure effect. *Journal of Vegetation Science*, **12**, 817-824.

Peschier, J. (1995) Characterisation of Topographic Surfaces on a Triangulated Irregular Network (TIN). In Geometrical Algorithms for Digital Elevation Models Utrecht University, The Netherlands.

Pierce, K.B., Lookingbill, T., & Urban, D. (2005) A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis. *Landscape Ecology*, **20**, 137-147.

Press, S.J. & Wilson, S. (1978) Choosing between Logistic Regression and Discriminant-Analysis. *Journal of the American Statistical Association*, **73**, 699-705.

Quinn, P.F., Beven, K.J., & Lamb, R. (1995) The Ln(a/Tan-Beta) Index - How to Calculate It and How to Use It within the Topmodel Framework. *Hydrological Processes*, **9**, 161-182.

Riano, D., Chuvieco, E., Salas, J., & Aguado, I. (2003) Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types (2003). *Ieee Transactions on Geoscience and Remote Sensing*, **41**, 1056-1061.

Rodriguez, E., Morris, C.S., Belz, J.E., Chapin, E.C., Martin, J.M., Daffer, W., & Hensley, S. (2005). An assessment of the SRTM topographic products. Jet Propulsion Laboratory, Pasadena, California.

Schmidt, F. & Persson, A. (2003) Comparison of DEM Data Capture and Topographic Wetness Indices. *Precision Agriculture*, **4**, 179-192.

Shepherd, J.D. & Dymond, J.R. (2003) Correcting satellite imagery for the variance of reflectance and illumination with topography. *International Journal of Remote Sensing*, **24**, 3503-3514.

Smith, A.P. & Young, T.P. (1987) Tropical Alpine Plant Ecology. *Annual Review of Ecology and Systematics*, **18**, 137-158.

SPSS (2003) Apache Software Foundation, Chicago.

SSTARS (2005) Multicollinearity in Logistic Regression. In http://www.uky.edu/ComputingCenter/SSTARS/MulticollinearityinLogisticRegression.htm. Center for Statistical Computing Support, Lexington, UK.

Stevens, G.C. & Fox, J.F. (1991) The Causes of Treeline. *Annual Review of Ecology and Systematics*, **22**, 177-191.

Tarboton, D.G. (1997) A New Method For The Determination Of Flow Directions And Upslope Areas In Grid Digital Elevation Models. *Water Resources Research*, **33**, 309-319.

UNEP-WCMC (2005) Sangay National Park, Ecuador. In http://sea.unep-wcmc.org/sites/wh/sangay.html

USGS (2003) Shuttle Radar Topography Mission. In http://edc.usgs.gov/srtm/index.html. EROS Data Center, Sioux Falls, SD, USA.

Virtanen, T., Mikkola, K., Nikula, A., Christensen, J.H., Mazhitova, G.G., Oberman, N.G., & Kuhry, P. (2004) Modeling the location of the forest line in northeast European Russia with remotely sensed vegetation and GIS-based climate and terrain data. *Arctic Antarctic and Alpine Research*, **36**, 314-322.

Walsh, S.J., Butler, D.R., Malanson, G.P., Crews-Meyer, K.A., Messina, J.P., & Xiao, N. (2003) Mapping, modeling, and visualization of the influences of geomorphic processes on the alpine treeline ecotone, Glacier National Park, MT, USA. *Geomorphology*, **53**, 129–145.

Wardle, P. (1985) New-Zealand Timberlines .2. a Study of Forest Limits in the Crow Valley near Arthurs Pass, Canterbury. *New Zealand Journal of Botany*, **23**, 235-261.

Wood, J. (2003) The Geomorphological Characterisation of Digital Elevation Models, City University, London.

Wu, J.G. (2004) Effects of changing scale on landscape pattern analysis: scaling relations. *Landscape Ecology*, **19**, 125-138.

Wuensch, K.L. (2005) Binary logistic regression with SPSS. In http://core.ecu.edu/psyc/wuenschk/MV/Multreg/Logistic-SPSS.doc. Department of Psychology, East Carolina University.

Yang, X., Chapman, G.A., Young, M.A., & Gray, J.M. (2005) Using Compound Topographic Index to Delineate Soil Landscape Facets from Digital Elevation Models for Comprehensive Coastal Assessment. In MODSIM 2005 International Congress on Modelling and Simulation (eds A. Zerger & R.M. Argent), pp. 170-176. Modelling and Simulation Society of Australia and New Zealand.