Estimating Vegetation Height and Canopy Cover from Remotely Sensed Data with Machine Learning

Daniela Stojanova[*,a], Panče Panov[*,b], Valentin Gjorgjioski[b], Andrej Kobler[a], Sašo Džeroski[b]

[a] *Slovenian Forestry Institute, Večna pot 2, SI-1000 Ljubljana, Slovenia*
[b] *Jožef Stefan Institute, Department of Knowledge Technologies, Jamova cesta 39, SI-1000 Ljubljana, Slovenia*

Abstract

High quality information on forest resources is important to forest ecosystem management. Traditional ground measurements are labor and resource intensive and at the same time expensive and time consuming. For most of the Slovenian forests, there is extensive ground-based information on forest properties of selected sample locations. However there is no continuous information of objectively measured vegetation height and canopy cover at appropriate resolution.

Currently, Light Detection And Ranging (LiDAR) technology provides detailed measurements of different forest properties because of its immediate generation of 3D data, its accuracy and acquisition flexibility. However, existing LiDAR sensors have limited spatial coverage and relatively high cost of acquisition. Satellite data, on the other hand, are low-cost and offer broader spatial coverage of generalized forest structure, but are not expected to provide accurate information about vegetation height.

Integration of LiDAR and satellite data promises to improve the measurement, mapping, and monitoring of forest properties. The primary objective of this study is to model the vegetation height and canopy cover in Slovenia by integrating LiDAR data, Landsat satellite data, and the use of machine learning techniques. This kind of integration uses the accuracy and precision of LiDAR data and the wide coverage of satellite data in order to generate cost effective realistic estimates of the vegetation height and canopy cover, and consequently generate continuous forest vegetation map products to be used in forest management and monitoring.

Several machine learning techniques are applied to this task: they are evaluated and their performance is compared by using statistical significance tests. Ensemble methods perform significantly better than single and multi-target regression trees and are further used for the generation of forest maps. Such maps are used for land-cover and land-use classification, as well as for monitoring and managing ongoing forest processes (like spontaneous afforestation, forest reduction and forest fires) that affect the stability of forest ecosystems.

Key words: remote sensing, LiDAR, Landsat, vegetation height, canopy cover, machine learning

[*] The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Email addresses: daniela.stojanova@gozdis.si (Daniela Stojanova), pance.panov@ijs.si (Panče Panov), valentin.gjorgjioski@ijs.si (Valentin Gjorgjioski), andrej.kobler@gozdis.si (Andrej Kobler), saso.dzeroski@ijs.si (Sašo Džeroski)

1. Introduction

In forest management and forestry decision-making there is a continuous need for high quality information on forest resources. The state of forest resources can be monitored by using visualizations of forest properties for a specific spatial region in the form of a map. Forest maps are an effective tool for detecting the state of forest resources and monitoring ongoing spatial processes in forested landscapes. Examples of such processes include the enlargement of forest area by spontaneous afforestation of abandoned agricultural land, and the vertical growth of trees and transitions between developmental stages of existing forest stands. These processes affect the stability of forest ecosystems, an ever more important property due to extreme weather conditions, hydrological stress and the appearance of new diseases and pests.

One of the most important forest properties are: vegetation height and canopy cover. Vegetation height is the height of the vegetation in a stand, relative to the ground. It is a function of the species composition, climate and site quality, and can be used for land cover classification or in conjunction with vegetation indices. If coupled with species composition and site quality information, vegetation height serves as an estimate of the stand age or the successional stages. Vegetation height is also a useful indicator of forest age and habitat quality. It is an important input variable for ecosystem and forest fire models, and is highly correlated with vegetation biomass and productivity. Biomass is the key component of the carbon circle (Skole and Tucker, 1993) and a surrogate for fuel loading estimation (Finney, 2004).

Forest canopy cover is defined as the percent cover of the tree canopy in a stand. It includes the cover from both trees and shrubs, but not herbal vegetation. Canopy cover describes the vertical projection of the tree canopy onto an imaginary horizontal surface representing the ground surface. Forest canopy cover is an ecologically very important forest property because it determines the occurrence and speed of forest regeneration. It is useful for distinguishing different plant and animal habitats, assessing forest floor microclimate, light conditions and estimating other forest variables (e.g., Leaf Area Index). Measurements of canopy cover are essential for silvicultural activities (Jennings et al., 1999).

Traditional ground-based field measurements of forest properties are made by using handheld equipment. These measurements are expensive, subjective, time consuming and labor intensive, as well as difficult to perform, especially in dense forests (Buckley et al., 1999). Due to these reasons, other methods of estimating forest properties for larger areas are often used, such as remote sensing.

Over the course of the past few decades, remote sensing[1] (RS) has been a valuable source of information in mapping and monitoring forest activities. Remote sensing involves collecting of spatially organized data and information about an area of interest by detecting and measuring signals composed of radiation, particles and fields emanating from objects located beyond the immediate neighborhood of the sensor devices (Franklin, 2001). In this way, it offers a potential for more efficient resource assessment.

Multi-spectral RS is often used to map structural metrics at moderate resolution and broader scale. Multi-spectral satellite imagery is well suited for capturing horizontally distributed (2D) conditions, strictures and changes (Wulder et al., 2008). However, it cannot capture the 3D forest structure directly and is easily influenced by topographical covers and weather conditions.

Light Detection And Ranging (LiDAR) technology, on the other hand, provides horizontal and vertical information (3D) at high spatial resolution and vertical accuracies. It good for

[1] Remote sensing. See also: `http://rst.gsfc.nasa.gov` (accessed February 11, 2010)

characterizing the vertical structure of vegetation, but has limited spatial coverage mostly due to pricing. By combining remotely sensed data, that describe the horizontal distribution of target phenomena, with LiDAR data, we can improve the measurement, mapping and monitoring of forest properties and provide means of characterizing forest canopy parameters and dynamics.

In this context, many papers have been recently published on the joint use of LiDAR and other active and passive sensors in forest properties estimation problems (Lefsky et al., 1999; Hyde et al., 2006; Maltamo et al., 2006). These studies perform estimation of the forest structure directly from LiDAR measurements and extend them, over limited areas, to spatially homogeneous spectral segments derived from the optical data sets. Medium resolution RS data, such as Landsat images, are relatively inexpensive to acquire over large areas (Franklin and Wulder, 2002), whereas LiDAR covers small areas, at a high cost per unit area (Lim et al., 2003). As a result, these two data types may be combined to generate estimates of vegetation heights and canopy cover over large areas at a reasonable cost (Hudak et al., 2002).

Latest studies (Wulder et al., 2008) of the integration of LiDAR and satellite data point out possible high correlations between different satellite images and forest properties (vegetation height and canopy cover). Hyde et al. (2006) compared the performance of step-wise linear regression models using waveform LiDAR, RaDAR, Landsat, Quickbird and InSAR in a statistical combination of structural information in an attempt to estimate the mean canopy height and biomass. The addition of Landsat ETM+ metrics significantly improved LiDAR estimates of large tree structure - the combination of all sensors is more accurate than using LiDAR alone, but only marginally better than the combination of LiDAR and Landsat ETM+.

Machine learning techniques, such as regression trees, artificial neural network and support vector machines have been widely used in many remote sensing forestry applications (Lefsky et al., 1999; Moghaddam et al., 2002; Wulder and Seeman, 2003). The typical machine learning task in all these studies is to learn a predictive model that uses a set of remote sensing observations with the aim of predicting the value of forest conditions or properties for unseen cases. The data input to the machine learning system consists of information extracted from different RS data sources, while the output of the system is a predictive model (or a set of predictive models called an ensemble) that describe the forest property.

The main objective of this study is to estimate the vegetation height and canopy cover from an integration of LiDAR and Landsat data in a diverse and unevenly distributed forest. This kind of integration uses the accuracy and precision of LiDAR data and the wide coverage of satellite data in order to generate cost effective realistic estimation of the forest properties over a geographically large area. The study area is located in the Kras region in western Slovenia, near the border with Italy. The input to the machine learning system are the independent explanatory variables generated from multi-temporal Landsat data and the target variables (representing forest properties that we want to model): The latter are estimated from the 3D LiDAR data and serve as a very good substitute for field-base sample plot measurements. The machine learning system outputs a predictive model of the forest property at hand, which is then used to generate forest vegetation maps that can be used in a variety of forest management applications.

Although forest vegetation maps can be generated with high precision and accuracy purely from LiDAR data, this seems impractical for the nearest future due to the very high cost of high resolution LiDAR data (in our case 4 EUR/hectare). On the other hand, the price of Landsat ETM+ data for a multi-temporal coverage is significantly lower (in our case it is free of charge). Using Landsat data as the main data source therefore ensures a very acceptable cost benefit ratio. On the other hand, LiDAR as used here for model calibration seems a very good substitute for field-based sample plot measurements of vegetation height and canopy cover, due to the even

higher costs of field measurements which can in some cases also be very difficult and imprecise.

In our preliminary work (Džeroski et al., 2006a,b; Taškova et al., 2006), we introduce the problem of prediction of forest parameters from Landsat and LiDAR data, and present preliminary results using a limited set of machine learning algorithms. The predictive models for estimating the vegetation height and canopy cover from LiDAR and Landsat data, using model and regression trees, pointed out a possible high correlation between satellite data and vegetation properties (Džeroski et al., 2006b). These results where enhanced by using additional machine learning techniques (bagging of model trees) in Taškova et al. (2006).

In this study, we significantly extend and upgrade the work presented in the preliminary work. Here we investigate the performance of a broader set of state-of-the-art machine learning techniques. We confirm the results from our preliminary work by systematically repeating the experiments using the same machine learning techniques. In addition, we apply other state-of-the-art machine learning techniques, i.e., ensemble methods that aim at improving the predictive performance of a given machine learning technique, using single (learning an ensemble for each target variable separately) as well as multi-target setting (learning an ensemble for all target variables together). We use a more carefully chosen experimental methodology that allows extensive comparisons of the predictive performances of all algorithms and perform statistical significance testing. Finally, we use the model with the best predictive power for generation of vegetation height and canopy cover maps of the Kras region of Slovenia and provide a more comprehensive discussion of the experimental results and the use of the map products .

The remainder of the paper is organized as follows. In Section 2, we first describe the data and the methodology used in this study. In Section 3, we then present the results of the modeling process. Next, in Section 4 we present a comparison of the models, discussion on the significance of the results and the map products. Finally, in Section 5 we outline our conclusions and discuss possible directions for further work.

2. Materials and Methods

2.1. Study Area

The study area measures 72226 hectares of the Kras region in western Slovenia, in the vicinity of the Adriatic sea, 5 *km* from the Gulf of Trieste. The local Gauss - Krueger coordinates of the study area are: $Min.Easting(X) = 389000$, $Max.Easting(X) = 433000$, $Min.Northing(Y) = 37000$ and $Max.Northing(Y) = 86000$.

The relief of the study area is rough with slopes ranging up to 60°, the average slope being 22°. The investigated area covers very diverse and not evenly distributed vegetation. The Kras region has about 40 different types of trees, which includes species such as: *Ostrya carpinifolia* (Hop-hornbeam), *Pinus nigra* (Black pine), *Quercus pubescens* (Downy Oak), *Fraxinus orneus* (South Europea Flowering Ash) and *Fagus syllvatica* (European Beech). In Figure 1 we present the map of Slovenia on which we mark the area recorded by LiDAR and the Kras region. The study area is encompassed with a black contour line, whereas the study area recorded with LiDAR is covered with black color. The white dots within the LiDAR area present parts not covered with vegetation i.e. denote settlements and were not included in the study.

2.2. Data Description

2.2.1. Data sources

Passive optical systems such as aerial photography and Landsat, as well as active systems like Radar and LiDAR, provide cost-effective methods of spatial data collection and measurements

of forest properties. The suitability a sensor type for a particular study depends on the scale of study and the nature of the observed objects or processes. In this study, we used the Landsat and LiDAR remote sensing techniques for estimating of the vegetation height and canopy cover.

Landsat. Landsat 7 Thematic Mapper Plus ETM+ [2] is the latest satellite of the Landsat Program designed to collect radiance data in 7 bands (channels) of reflected energy and one band of emitted energy. A well calibrated ETM+ enables one to convert the raw solar energy collected by the sensor to absolute units of radiance. The eight bands of ETM+ data are used to discriminate between Earth surface materials through the development of spectral signatures. Thus, a multi-spectral data set having both high (30 m) and medium to coarse (250 m-1000 m) spatial resolution is acquired on a global basis repetitively and under nearly identical atmospheric and plant physiological conditions. The panchromatic band has spatial resolution of 15 m, while the thermal infrared (TIR) channel has a resolution of 60 m .

LiDAR. Airborne laser scanning (ALS), also termed airborne LiDAR (Light Detection And Ranging) is an optical remote sensing technology that measures properties of scattered light to find range and/or other information of a distant target. The laser emits a light pulse which is scattered (reflected) from the object back to the sensor. By measuring the round trip time of an emitted laser pulse from the sensor to a reflecting surface and back again, the distance from the sensor to the surface is determined.

LiDAR is one of the most promising remote sensing techniques for detailed measurements of forest properties because of its immediate generation of 3D data, self-georeferencing, high spatial resolution (typically 0.5-5 $points/m$, positional error 10-20 cm), accuracy (raging from 15-20 cm Root Mean Square Error (RMSE) vertically and 20-30 cm horizontally) and acquisition flexibility [3]. It enables detailed measurements and making of maps with quality comparable to the most passive or active systems. It penetrates through the vegetation layer to the bare ground, enabling structural rendering of vegetation and providing 3D data about objects.

With LiDAR, we can directly define the third dimension of forest layers and the relief under the forest. It is a good source for generation of digital relief models (DEM) and topographical analysis, especially for forested areas, where classical aerophotogrametrical techniques do not give satisfactory accuracy. LiDAR can be used for mapping forest stands, individual tree canopy detection, etc.

2.2.2. Data description and generation of the dataset

The data used in this study consists of multi-spectral multi-temporal Landsat satellite images and 3D LiDAR recordings of the study area. From the Landsat data, we extracted the explanatory variables, while the LiDAR data was used to extract the target variables (forest properties) used in the process of learning the predictive model. The spatial unit of analysis was a 25 $m \times 25$ m square.

[2] Landsat. See also: `http://www.trfic.msu.edu/data_portal/Landsat7doc/landsatch5.html` (accessed February 11, 2010)

[3] Instrument technical details. See also: `http://arsf.nerc.ac.uk/instruments/altm.asp` (accessed August 18, 2008)

Landsat data description. Multi-spectral Landsat ETM+ data were acquired on August 3rd, 2001, May 18th, 2002, November 10th, 2002, and March 18th, 2003, thus capturing the main phenological stages of forest vegetation in the area. In Figure 2 we show a part of a Landsat ETM+ band 3' image, that covers the area recorded with LiDAR, obtained on November 10th, 2002. The Landsat imagery was first geometrically corrected by orthorectification. Image segmentation was then applied. The commercially available eCognition image analysis software, version 2.1 (Definiens Imaging, Munich, Germany) was used for the image segmentation. The software uses a patented procedure for multi-resolution segmentation to extract image objects, exploiting both spatial and spectral information to create objects from image data. The segmentations are typically visually appealing, although the users need to interactively select a useful segmentation level through trial and error (Hay et al., 2003).

The typical result of image segmentation is extraction of large homogeneous image objects (e.g., meadow), small homogeneity image objects (e.g., forest stand) and small homogeneity objects embedded in a high contrast, especially for data such as Landsat imagery. Each of the four Landsat images was segmented at two levels of spatial detail in order to get realistic object based information that correspond to the real world objects and later serve as information carrier and building block for further analysis. The average image segment sizes were 4 *ha* for the fine segmentation and 20 *ha* for the coarse segmentation. Image segmentation is illustrated in Figure 3 and it represents a segmentation of the Landsat image presented in Figure 2. It has been derived as a result of fine image segmentation of the third Landsat channel. The objects are given with different color in order to be distinguishable among each other (the number of objects is around 45500).

Explanatory variables. In order to represent and display remote sensed data, we employ basic statistic measures like band mean value, standard deviation and others (Jensen, 2004). The statistic measures can be used further in the analysis of the data directly or indirectly. The link between remote sensing and statistics is strong; clearly, remote sensing can be considered a multivariate problem (Kershaw, 1987) and probabilistic methods constitute one of the most powerful approaches to the analysis of multivariate problems.

Therefore, we generate our explanatory variables from Landsat imagery data based on statistical information for each band. Based on the data within each image segment, four statistic measures (minimum reflectance, maximum reflectance, average reflectance, standard deviation of reflectance) were computed for each date, for each segmentation level, and for each of the Landsat image channels (2, 3, 4, 5, 7). Using different segmentation levels, for each example, we take into account two different kind of neighborhood (narrow and broader). The information about the narrow neighborhood is included with the fine image segmentation level and the broader one is included with the coarse image segmentation level. In this way, we obtain 160 explanatory variables to be used in the predictive modeling. As the borders of individual segments were not identical between dates and segmentation levels, values of the 160 variables were attributed back to individual image pixels, each with dimension 25 $m \times 25$ m.

LiDAR data description. An east-west transect measuring 2 $km \times 20$ km (highlighted in black in Figure 1) across a representative part of the Kras region was flown over by LiDAR, in the spring of 2005. The equipment included Optech ALTM 3100 LiDAR flown on a Eurocopter EC-120 B "Colibri" helicopter. The device collects 33 000 laser observations per second in standard operating mode, measuring height, first, intermediate, only and last returns, angle, radian and intensity data. From an operating altitude of 1000 m, the resulting height data has an absolute

root mean squared error better than ±15 cm. The average point cloud density of the LiDAR dataset was 7.5 $points/m^2$, thus 4687.5 discrete 3D LiDAR returns were contained on average in each 25 $m \times 25$ m square .

Target variables. The target variables were computed from the LiDAR data, at the level of 25 m $\times$ 25 m squares corresponding to Landsat pixels. The vegetation height (H) for each square (or Landsat pixel) was computed by averaging the heights of the LiDAR-based normalized digital surface model (nDSM) within the 25 $m \times 25$ m square. A nDSM is a high resolution raster map showing the relative height of vegetation above the bare ground. Our nDSM had a horizontal resolution of 1 m^2 and was computed using the REIN (REpetitive INterpolation) algorithm for calculation of a Digital Terrain Model (DTM) (Kobler et al., 2007). The REIN algorithm was developed for generating DTMs under forest cover in steep terrain using dense LiDAR data (≥ 5 $points/m^2$): In such conditions, other filtering algorithms typically have problems distinguishing between ground returns and off-ground points reflected in the vegetation. A field validation of the nDSM on a sample of 120 trees confirmed a vertical RMS error of 0.36 m and a vertical bias of -0.71 m.

The canopy cover (CC) within this study is defined as the percentage of bare ground within 25 $m \times 25$ m (or a Landsat pixel), covered by the vertical projection of the overlying vegetation, higher than 1 m. The canopy cover for each Landsat pixel was computed as the ratio of the heights of the LiDAR-based normalized digital surface model (nDSM) that exceeded 1 m relative height difference between the bare ground of the digital terrain model and the surface of the Landsat pixel. The canopy cover for each 25 m square was computed as the percentage of vegetation within a pixel. The values of the canopy cover are in the interval 0-100%.

2.3. Machine learning methodology

Predictive modeling is a machine learning task concerned with predicting the value of one or more dependent variables (classes, targets) from the values of independent variables (explanatory variables). If the target variable is continuous, the task at hand is called regression. If the target is discrete (it has a finite set of nominal values), the task at hand is called classification. The tasks of classification and regression are the two most commonly addressed predictive modeling tasks in machine learning.

In predictive modeling, a set of data records is taken as input to a predictive modeling algorithm, and a predictive model (or set of predictive models called an ensemble) is generated as an output. This model can then be used to predict values of the target variable for new data. If we are predicting a value of a single target variable, then we have a single-target prediction task. In the case when we predict the values of several target variables simultaneously with one model, we have a multi-target prediction task.

In this study, the machine learning task is to learn a predictive model (or a set of models) for predicting vegetation height and canopy cover from an integration of LiDAR and Landsat data. This is a multi-target prediction task. The target variables are derived from the LiDAR data and the explanatory variables are extracted from the Landsat images.

2.3.1. Single-target prediction: decision, regression and model trees

Decision tree learning (Quinlan, 1986) is one of the most widely used methods for inductive learning. A decision tree is a hierarchical structure, where the internal nodes contain tests on the descriptive variables. Each branch of an internal test corresponds to an outcome of the test, and the prediction for the value of the target variable is stored in a leaf. To obtain a prediction for a

new data record, the record is sorted down the tree, starting from the root (the top-most node of the tree). For each internal node that is encountered on the path, the test is stored in the applied node. Depending on the outcome of the test, the path continues along the corresponding branch. The resulting prediction of the tree is taken from the leaf at the end of the path.

A decision tree is usually constructed with a recursive partitioning algorithm from a training set of records. The algorithm is known as Top-Down Induction of Decision Trees (TDIDT). The records include measured values of the descriptive and the target attributes. The tests in the internal nodes of the tree refer to the descriptive,while the predicted values in the leaves refer to the target attributes.

The TDIDT algorithm starts by selecting a test for the root node. Based on this test, the training set is partitioned into subsets according to the test outcome. In the case of binary trees, the training set is split into two subsets: one containing the records for which the test succeeds (typically the left subtree) and the other containing the records for which the test fails (typically the right subtree). This procedure is recursively repeated to construct the subtrees.

The partitioning process stops when a stopping criterion is satisfied (e.g., the number of records in the induced subsets is smaller than some predefined value; the length of the path from the root to the current subset exceeds some predefined value, etc.). In that case, the predicted value is calculated and stored in a leaf. The predicted value is the mean value of the target variable calculated over the records that are sorted into the leaf.

One of the most important steps in the tree induction algorithm is the test selection procedure. For each node a test is selected by using a heuristic function computed on the training data. The goal of the heuristic is to guide the algorithm toward smaller trees with good predictive performance.

Regression trees are decision trees that predict the value of a numeric target attribute (Breiman et al., 1984). Each leaf of a regression tree contains a constant value as a prediction for the target variable, as regression trees represent piece-wise constant functions. If the leaf contains a linear regression model that predicts the target value of examples that reach the leaf, the decision tree in question is called a model tree (Quinlan, 1992). Model trees have advantages over regression trees in both compactness and prediction accuracy, and the ability to exploit local linearity in the data. Another advantage over regression trees is that model trees can extrapolate the predicted value outside the range observed in the training cases. In this paper, we use $M5'$ regression and model tree algorithm implementation from the WEKA environment (Witten and Frank, 2005).

2.3.2. Multi-target prediction: multi target regression trees

Multi-target regression trees (Blockeel, 1998; Struyf and Džeroski, 2006) are a generalization of regression trees for the prediction of several numeric target variables simultaneously. The leaves of a multi-target regression tree store a vector of numeric values, instead of storing a single numeric value. Each component of this vector is a prediction for one of the target attributes. The components of the prediction vector are the means of the target variables calculated over the records that are stored in the leaf. The main advantages of multi target regression trees (over building a separate model for each target) are: (1) a multi-objective model is smaller than the total size of the individual models for all target variables, and (2) such a multi-objective model explicates dependencies between the different target variables.

In this paper, we use the CLUS (Blockeel and Struyf, 2002; Struyf and Džeroski, 2006) system for constructing (multi-target) regression trees. The heuristic used for selecting the attribute tests (that define the internal nodes) in this algorithm is the intra-cluster variance summed over

the subsets induced by the test. The variance function is standardized so that the relative contribution of the different targets to the heuristic score is equal. Lower intra-subset variance results in predictions that are more accurate.

2.3.3. Ensembles

An ensemble method constructs a set of predictive models called an ensemble (Dietterich, 2000). An ensemble gives a prediction for a new data record by combining the predictions of the individual models for that data record. For regression tasks, the final prediction can be obtained by averaging the output predictions of the models in the ensemble. The learning of ensembles consists of two steps. In the first step, we have to learn the base models that make up the ensemble. In the second step, we have to figure out how to combine these models (or their predictions) into a single coherent model (or prediction).

When learning base models it makes sense to learn models that are accurate and diverse (Hansen and Salamon, 1990). Accurate models perform better than random guessing on new examples, and diverse models make different prediction errors on new examples. The diversity in an ensemble can be introduced in different ways: by manipulating the training set (e.g., bootstrap sampling, change of weights of the data instances) or by manipulating the learning algorithm used to obtain the base models (e.g., introducing random elements in the algorithm).

Ensemble methods aim at improving the predictive performance of a given machine learning technique. They aim to improve the predictive performance of their base classifier when used in a single target setting (learn an ensemble for each target attribute separately) (Breiman, 1996, 2001). In (Kocev et al., 2007), it is shown that this applies also for the multi-target setting (learn one ensemble for all target attributes). In addition, the ensembles for multi-target prediction should be preferred because they are faster to learn. In this work, we use bagging and random forests, the two most widely used ensemble methods to produce ensembles of regression trees and multi-target regression trees.

Bagging. Bagging (Breiman, 1996) is an ensemble method that constructs the different base models by making bootstrap replicates of the training set and using them to build the individual models. Each bootstrap sample is obtained by randomly sampling training instances, with replacement, from the original training set. The bootstrap sample and the training set have an equal number of instances. Bagging can give substantial gains in predictive performance, when applied to an unstable learner (i.e., a learner for which small changes in the training set result in large changes in the predictions), such as classification and regression tree learners.

Random forest. A random forest (Breiman, 2001) is an ensemble of trees, where the diversity among the individual trees is obtained from two sources: (1) by using bootstrap sampling and (2) randomization of the selection step of the TDIDT algorithm. At each node in the decision tree, a random subset of the input attributes is taken and the best split is selected from this subset. The size of the random subset is given by a function of the number of descriptive attributes. Prediction is made by aggregation (majority vote for classification or averaging for regression) of the predictions of the individual models in the ensemble.

3. Results

3.1. Experimental design

3.1.1. Dataset

The dataset consists of 160 explanatory variables and 2 target variables. The explanatory variables are derived from Landsat data for two levels of image segmentation, as explained in Section 2. The target variables are: vegetation height (H) and canopy cover (CC), derived from LiDAR data. There are 64000 examples of which 60607 describe the vegetation outside a settlement and are used in the process of learning.

3.1.2. The learning algorithms

In this study, one of the objectives is to study the predictive performance of state-of-the art machine learning algorithm, for the task of prediction of vegetation height and canopy cover. The problem of prediction of forest properties inherently represents a multi-target learning problem: it can be solved by using algorithms that build a single-target model for each forest property separately or by using algorithms that build a multi-target model for both forest properties at the same time. Another dimension of comparison of the predictive performance is using single models or ensemble of models. In this study, we investigate this dimension by performing experiments for single-model prediction and state-of-the-art ensemble prediction (e.g., bagging and random forests) both in the single-target and multi-target setting.

We use implementations of the state-of-the-art algorithms from two open source machine learning systems: WEKA (Witten and Frank, 2005) and CLUS[4] (Blockeel and Struyf, 2002; Struyf and Džeroski, 2006). In total, we performed experiments using 9 different algorithms. First, we performed experiments using algorithms that have a single model as an output. We used the implementations of regression tree (wRT) and model tree (wMT) algorithm in the WEKA system and single-target (STRT) and multi-target regression trees (MTRT) implemented in the CLUS system. Next, we performed experiments using ensemble learning algorithms that produce a set of models. In this case, we used the implementations of the bagging method from WEKA using model trees as base-level learners (wBagMT), and bagging and random forests of CLUS regression trees (as base learners) in the CLUS system both in the single-target (BagSTRT and RFSTRT) and multi-target setting (BagMTRT and RFMTRT).

The experiments were performed by using the default parameter settings for all the algorithms. Single-target regression trees and multi-target regression trees from the CLUS system are built with the default heuristic (intra-cluster variance) and default pruning method (M5 pruning). The minimal number of examples for the method to form a leaf was 4 examples. The settings for ensembles include the default pruning method, the number of variables in variable selection for random forest was set to 5 variables (calculated using the suggestion by Breiman (2001)), the default ensemble size of 10 and the default voting type for regression (the mean value).

3.1.3. Evaluation and comparison

Evaluation of the models was preformed using the standard 10 fold cross-validation evaluation method. All the algorithms were evaluated on the same folds, in order to allow comparison of the results and statistical significance testing. We use two regression evaluation measures

[4]The system is available at `http://www.cs.kuleuven.be/~dtai/clus/` (accessed August 18, 2008)

to estimate and discuss the predictive performance of the models: correlation and root mean squared error. Correlation (Corr) indicates the strength and direction of a linear relationship between two random variables and is usually expressed through the Pearson correlation coefficient. Root mean squared error (RMSE) is a frequently-used measure of the differences between values predicted by a model of an estimator and the target values actually observed.

To compare the performance of the different algorithms, we use the corrected Friedman test (Friedman, 1940; Iman and Davenport, 1980). The evaluation measure for each fold of the cross-validation represents a data point for the statistical test. The test is performed on each target variable (H and CC) separate for each evaluation measure (Corr and RMSE).

The Friedman nonparametric test first ranks the algorithms for each dataset (fold), the best performing algorithm getting the rank of 1. It then compares the average ranks of the algorithms across datasets (folds). The null-hypothesis, which states that all the algorithms are equivalent and so their ranks should be equal.

If the null-hypothesis is rejected, we can proceed with a post-hoc test. The Nemenyi test Nemenyi (1963) is used when in our case, since all classifiers are compared to each other. The performance of two classifiers is significantly different if the corresponding average ranks differ by at least the critical difference CD. The results of this test are visualized by using the average rank diagrams on which the critical distance is also depicted (Demšar, 2006). We consider the differences in performance significant if the standard p-value is below the threshold of 0.05.

3.2. Results - predictive performance

Here, we present the predictive performance of the obtained models in terms of two evaluation measures (Corr and RMSE) for both target variables. The results, presented in Tables 1 and 2, are represented with the corresponding confidence intervals, to show the stability of the used algorithms. We can note that the confidence intervals in both tables are small, due to the size of the dataset (60607 examples). In Tables 1*a* and 2*a* we list the performance for algorithms that produce single models as output, and in Tables 1*b* and 2*b* we list the performance of ensemble algorithms.

To check whether the differences in performances are statistically significant, we used the corrected Friedman test for multiple hypothesis testing. To detect which algorithms perform significantly better or worse than others, we used the Nemenyi post hoc test. The results of this procedure are presented in the form of average rank diagrams in Figure 4, for each target variable and each evaluation measure. The ranks are depicted on the axis in such a manner that the best ranking algorithms are at the right-most side of the diagram. The critical difference (CD) interval, for a significance level of 0.05, is computed by the Nemenyi test and is plotted in the upper left corner; algorithms whose average rank difference is larger than this critical difference can be considered significantly different with 95 % probability. The algorithms that do not differ significantly are connected with a line.

The Nemenyi test shows (Figure 4*a* and 4*b*) that the best performing algorithms are ensemble methods and in particular random forests of multi-target regression trees (RFMTRT), while the worst performing algorithms are single-model algorithms. The test shows that the performance of the ensemble methods, in terms of correlation coefficient, is significantly better than the one of single-model methods. If we compare the multi-target methods, we can see that random forests of multi-target regression trees perform statistically better than individual multi-target regression trees: in the case of bagging, the difference is not statistically significant. Similar conclusions can be drawn if instead of the results for correlation we consider the results for RMSE (see Figures

4*c* and 4*d*). In general, RFMTRT constructed from the CLUS system perform significantly better than any of the individual trees. The only exception to this is the RMSE for canopy cover, where multi target regression trees (MTRT) have the same rank as RFMTRT.

3.3. Results - Maps of vegetation height and canopy cover

The second objective of our work is to produce maps of vegetation height and canopy cover using the predictive models obtained in the study. For that purpose, we used RFMTRT, which is the best performing method according to predictive performance, to generate maps. This model was built using the entire dataset of 60607 examples, from the representative part of the Kras region (containing variety of different vegetation) for which we have both Landsat and LiDAR data available. Next, we translated the RFMTRT model into functions in the PYTHON [5] programming language, that were later on used in the GIS (Geographical Information System) system to visualize the predictions in the form of a map. Finally, we generated maps of vegetation height (see Figure 5) and canopy cover (see Figure 6) by applying the PYTHON functions to the whole Kras region, thus extrapolating the predictions of the model built on the smaller representative part of the region using Landsat data available for the whole region.

4. Discussion

In this study, we compare several machine learning methods on the task of estimating vegetation height and canopy cover by using LiDAR and Landsat data. To this end, we redesigned the experiments from the first two preliminary studies (Džeroski et al., 2006b; Taškova et al., 2006). We tested additional machine learning methods in order to improve the accuracy of the predictive models. Beside single and multi-target regression trees used in the previous studies, we also use single and multi target ensemble methods.

The best results are obtained using the RFMTRT algorithm, random forests of multi-target regression trees. Ensemble methods improve the accuracy of the predictive models. Moreover, the ensembles for multi-target prediction should be preferred because they are faster to learn and predict more than one variable at the same time.

All ensemble methods perform better than the single model algorithms (wMT, wRT, STRT and MTRT) used. An exception is the performance in terms of the RMSE for canopy cover where MTRT have the same performance as RFMTRT. The average rank diagram show that random forests created by CLUS system perform best in all four cases (see Figure 4). The difference of the performance between ensembles of different types of trees is insignificant.

The results from this study are better than results presented in our preliminary work. Džeroski et al. (2006b) reported a correlation of 0.885 and RMSE=2.25 *m* for vegetation height and a correlation of 0.861 and RMSE=0.17 for canopy cover: These were achieved by using model trees. Taškova et al. (2006) reported a correlation of 0.902 and RMSE=2.19 *m* for vegetation height and a correlation of 0.882 and RMSE=0.238 for canopy cover: These were achieved by using bagging of model trees. The accuracy of the predictive models is improved by using ensemble methods. In this more general study, we obtained higher correlation coefficients and lower error rates. The average error rate (RMSE) of the best models is 2.05 *m* for the vegetation height and 14% for the canopy cover, whereas the corresponding correlation coefficients are 0.91 and 0.88.

[5] `http://www.python.org/` (accessed on August 18, 2008)

The investigated study area covers very diverse and not evenly distributed vegetation. It was selected by taking into account the diversity and the distribution of the many different vegetation types present in the Kras region. The Kras region has about 40 different types of trees, which includes species such as: *Ostrya carpinifolia* (Hop-hornbeam), *Pinus nigra* (Black pine), *Quercus pubescens* (Downy Oak), *Fraxinus orneus* (South Europea Flowering Ash) and *Fagus syllvatica* (European Beech). The models build using the methodology described in this paper can also serve for estimation of the vegetation height and canopy cover in other study areas with similar vegetation species. The different vegetation types have different influence on the structure and the accuracy of the model. The different combination of vegetation species will decrease (in most of the cases) the accuracy of the predictions of the model. In case of regions with very diverse vegetation it is preferable to divide the region into smaller subregions and perform modeling in each subregion separately. In addition, special attention when modeling the vegetation properties needs to be focused on the relief of the area.

The generated maps represent a rough, but continuous estimates of the vegetation height and canopy cover over a large spatial area. The precision of the derived maps is lower than the precision of the field measurements done on smaller plots or individual trees within the study area (see field validation of the nDSM in Section 2.2.2). Therefore, these maps cannot be used for determination of the growing stock or other individual tree estimates, but can be useful when coverage of a grater spatial area is required.

Such maps can be used as an input for advanced systems such as GIS to improve their planning, managing and monitoring capabilities, in performing a variety of tasks such as land cover mapping, land cover classification, land use mapping, land use classification, change detection and many other forestry, ecological, geological and military applications. Moreover, the maps can be used for monitoring and managing a variety of ongoing processes in the forest ecosystems that involve enlargement of forest areas by spontaneous afforestation of abandoned agricultural land, forest area reduction, urban rapprochement, as well as vertical growth and gradual closing of canopy cover of existing forest stands. These maps can be used in the process of monitoring the forest biomass accumulation and CO_2 sink in the Kyoto framework [6]. Furthermore they can be used in estimating the risk of forest fire outbreaks.

In addition, these maps can also serve for temporal comparisons. Finally, due to their spatial continuity (unlike the discrete sampling layout of current forest monitoring schemes) potential applications also include the study of forest habitats and transitional agricultural-forest habitats, visual landscape assessments, land use suitability analysis, visibility analysis for cell phone networks etc. The methodology used in this study integrates remote sensing, forestry and machine learning techniques and can be a powerful tool for diverse mapping and modeling applications in the future.

5. Conclusions

In this study, we focus on the estimation of forest properties (forest vegetation height and canopy cover) from remotely sensed data over a large geographical area (the study area measures 72226 hectares of the Kras region in western Slovenia in the vicinity of the Adriatic sea), by integrating LiDAR and Landsat satellite data and generating predictive models of forest properties. We use machine learning methods for predictive modeling and apply a set of state-of-the-art machine learning techniques. To model the forest properties we focused on two dimensions:

[6] Kyoto protocol: `http://unfccc.int/resource/docs/convkp/kpeng.html`, (accessed August 18, 2008)

modeling the parameters with individual models or ensembles (single model prediction and ensemble prediction) and modeling the target properties separately or simultaneously (single target and multi target prediction). The results show the advantages of multi target over single target regression, as multi target models have a smaller size and are faster to learn and apply, and the advantage of ensemble prediction over single model prediction in terms of predictive performance.

Several contributions are presented in this study. First, we use state-of-the-art machine learning methodology to model forest properties, in contrast to the simple statistical methods and linear regression used in similar studies (Hyde et al., 2006). Second, we achieved better results in terms of higher correlation coefficients and lower RMSE errors compared to the results obtained in our preliminary work (Džeroski et al., 2006b; Taškova et al., 2006). Also, we perform modeling of the forest properties in diverse forests, as opposed to modeling of homogeneous forests. Next, we use multi-temporal multi-spectral Landsat data, obtained in different vegetation seasons, instead of mono-temporal data used in similar studies. Finally, we use the accurate and precise LiDAR data to learn models for the representative part of a region and then we extrapolate the predictions on a larger area using less expensive remote sensing Landsat data.

The derived models represent a key piece of infrastructure required in support of sustainable forest management. They serve to generate forest vegetation map products for a large geographical area. Although such maps could be generated with exceeding precision and accuracy purely from LiDAR data, this seems impractical for the foreseeable future due to the very high cost of high resolution LiDAR data. Using Landsat data as the main data source therefore ensures a very acceptable cost benefit ratio. Moreover, using LiDAR for model calibration seems a very good replacement for sample plot field measurements of vegetation height and canopy cover, due to the even higher costs and difficulty or imprecision of the field measurements.

In future work, we first plan to investigate different image segmentation algorithms and to see what is the influence of segmentation on the overall predictive performance. Moreover, we would like to use other preprocessing methods and techniques and combine them with domain-based knowledge (e.g., image clustering, geo-ontologies). Second, we want to incorporate the spatial correlation and the spatial autocorrelation in the predictive models. Finally, we plan to expand the forest maps to broader areas (i.e., country level). We will evaluate the predictions of the machine learning models on different study areas and explore the influence of diverse vegetation and land cover types on the accuracy of the results.

Acknowledgment

The acquisition of the LiDAR and satellite data was done within the project "Processing LiDAR data (Development and usage of algorithms for mapping and estimation of forest stand biomass and structure using LiDAR and digital multispectral imagery) No. L2-6575, 2004-2007" funded by the Ministry of Education, Science and Sport of Republic of Slovenia.

References

Blockeel, H., 1998. Top-down induction of first order logical decision trees. Ph.D. thesis, Katholieke Universiteit Leuven, Belgium.

Blockeel, H., Struyf, J., 2002. Efficient algorithms for decision tree cross-validation. Journal of Machine Learning Research 3 (Dec), 621–650.

Breiman, L., 1996. Bagging predictors. Machine learning 24 (2), 123–140.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A., 1984. Classification and Regression Trees. The Wadsworth statistics/probability series. Chapman & Hall/CRC.

Buckley, D. S., Isebrands, J., Sharik, T. L., 1999. Practical field methods of estimating canopy cover, PAR, and LAI in Michigan Oak and pine stands. Northern Journal of Applied Forestry 16 (1), 25–32.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30.

Dietterich, T. G., 2000. Ensemble methods in machine learning. In: MCS '00: Proceedings of the First International Workshop on Multiple Classifier Systems. Springer-Verlag, London, UK, pp. 1–15.

Džeroski, S., Kobler, A., Gjorgjioski, V., Panov, P., 2006a. Predicting forest stand height and canopy cover from landsat and lidar data using data mining techniques. Poster presentation at Second NASA Data Mining Workshop: Issues and Applications in Earth Science, May 23-24, Pasadena, PA. 2006.

Džeroski, S., Kobler, A., Gjorgjioski, V., Panov, P., September 2006b. Using decision trees to predict forest stand height and canopy cover from Landsat and LiDAR data. In: Tochtermann, K., Scharl, A. (Eds.), Managing environmental knowledge : EnviroInfo 2006 : proceedings of the 20th International Conference on Informatics for Environmental Protection. Aachen: Shaker Verlag, Graz, Austria, pp. 125–133.

Finney, M. A., 2004. FARSITE: Fire Area Simulator-model development and evaluation. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station., Res. Pap. RMRS-RP-4, Ogden, UT.

Franklin, S., Wulder, M., 2002. Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. Progress in Physical Geography 26 (2), 173–205.

Franklin, S. E., 2001. Remote Sensing for Sustainable Forest Management. Lewis Publishers.

Friedman, M., 1940. A comparison of alternative tests of significance for the problem of m rankings. The Annals of Mathematical Statistics 11 (1), 86–92.

Hansen, L., Salamon, P., 1990. Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (10), 993–1001.

Hay, G. J., Blaschke, T., Marceau, D. J., Bouchard, A., April 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. ISPRS Journal of Photogrammetry and Remote Sensing 57 (5), 327–345.

Hudak, A., Lefsky, M., Cohen, W., Berterretche, M., 2002. Integration of LiDAR and Landsat ETM+ data for estimating and mapping forest canopy height. Remote Sensing of Environment 82 (2), 397–416.

Hyde, P., Dubayah, R., Walker, W., Blair, J. B., Hofton, M., Hunsaker, C., 2006. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. Remote Sensing of Environment 102 (1-2), 63 – 73.

Iman, R., Davenport, J., 1980. Approximations of the critical region of the Friedman statistic. Comm. Stat. Theor. Meth. A9 (6), 571–595.

Jennings, S., Brown, N., Sheil, D., 1999. Assessing forest canopies and understory illumination: canopy closure, canopy cover and other measures. Forestry 72 (1), 59–74.

Jensen, J. R., 2004. Introductory Digital Image Processing (3rd Edition). Prentice Hall Series in Geographic Information Science. Prentice Hall.

Kershaw, C., 1987. Discrimination problems for satellite images. International Journal of Remote Sensing 8 (9), 1377–1383.

Kobler, A., Pfeifer, N., Ogrinc, P., Todorovski, L., Oštir, K., Džeroski, S., 2007. Repetitive interpolation: A robust algorithm for DTM generation from Aerial Laser Scanner Data in forested terrain. Remote Sensing of Environment 108 (1), 9–23.

Kocev, D., Vens, C., Struyf, J., Džeroski, S., 2007. Ensembles of multi-objective decision trees. In: Machine Learning: ECML 2007, 18th European Conference on Machine Learning, Proceedings. Vol. 4701 of Lecture Notes in Computer Science. Springer, pp. 624–631.

Lefsky, M. A., Cohen, W. B., Hudak, A., Acker, S. A., Ohmann, J. L., 1999. Integration of LIDAR, Landsat ETM+ and forest inventory data for regional forest mapping. In: Proceedings of the ISPRS Workshop Mapping Surface structure and topography by airborne and spaceborne lasers. Vol. XXXII-3/W14 of International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., Flood, M., 2003. LiDAR remote sensing of forest structure. Progress in Physical geography 27 (1), 88–106.

Maltamo, M., Malinen, J., Packaln, P., Suvanto, A., Kangas, J., 2006. Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. Canadian Journal of Remote Sensing 36 (2), 426–436.

Moghaddam, M., Dungan, J., Acker, S., 2002. Forest variable estimation from fusion of SAR and multispectral optical data. IEEE Transactions on Geoscience and Remote Sensing 40 (10), 2176–2187.

Nemenyi, P., 1963. Distribution-free multiple comparisons. Ph.D. thesis, Princeton University, Princeton, NY, USA.

Quinlan, J. R., 1986. Induction of decision trees. Machine Learning 1 (1), 81–106.

Quinlan, J. R., 1992. Learning with continuous classes. In: Proceedings of the 5th Australian Joint Conference on

649 Artificial Intelligence. World Scientific, pp. 343–348.
650 Skole, D., Tucker, C., 1993. Tropical Deforestation and Habitat Fragmentation in the Amazon: Satellite Data from 1978
651 to 1988. Science 260 (5116), 1905–1910.
652 Struyf, J., Džeroski, S., 2006. Constraint based induction of multi-objective regression trees. In: Knowledge Discovery
653 in Inductive Databases, 4th International Workshop, KDID'05, Revised, Selected and Inductive Papers. Vol. 3933 of
654 Lecture Notes in Computer Science. Springer, pp. 222–233.
655 Taškova, K., Panov, P., Kobler, A., Džeroski, S., Stojanova, D., 2006. Predicting forest stand properties from satellite
656 images with different data mining techniques. In: Proceedings of the 9th International Multiconference Information
657 Society IS 2006, 9th-14th October 2006, Ljubljana, Slovenia. pp. 259–262.
658 Witten, I., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition. Morgan
659 Kaufmann, San Francisco.
660 Wulder, A., Seeman, D., 2003. Forest inventory height update through the integration of Lidar data with segmented
661 Landsat imagery. Canadian Journal of Remote Sensing 29 (5), 536–543.
662 Wulder, M., White, J., Hay, G., Castilla, G., 2008. Object-Based Image Analysis. Lecture Notes in Geoinformation
663 and Cartography. Springer Berlin Heidelberg, Ch. Pixels to objects to information: Spatial context to aid in forest
664 characterization with remote sensing, pp. 345–363.

Table 1: Comparison of correlation coefficients of the predictive models for both target variables: *a*) Single model algorithms (**wRT** - WEKA Regression Tree; **wMT** - WEKA Model Tree; **STRT** - CLUS Single Target Regression Tree; **MTRT** - CLUS Multi target Regression Tree); *b*) Ensemble algorithms (**wBagMT** - WEKA Bagof Model Trees; **BagSTRT** - CLUS Bag of STRTs; **RFSTRT** - CLUS Random Forest of STRTs; **RFMTRT** - CLUS Random Forest of MTRTs)

a) **Single model algorithms**

	Single target			Multi-target
Target	wRT	wMT	STRT	MTRT
H	0.876 ± 0.004	0.884 ± 0.004	0.874 ± 0.003	0.880 ± 0.015
CC	0.858 ± 0.002	0.863 ± 0.004	0.851 ± 0.003	0.852 ± 0.013

b)**Ensemble algorithms**

	Single target			Multi-target	
Target	wBagMT	BagSTRT	RFSTRT	BagMTRT	RFMTRT
H	0.902 ± 0.004	0.904 ± 0.003	0.906 ± 0.002	0.904 ± 0.002	0.906 ± 0.002
CC	0.883 ± 0.002	0.880 ± 0.003	0.883 ± 0.002	0.880 ± 0.002	0.883 ± 0.002

Table 2: Comparison of RMSE of the predictive models for both target variables: *a*) Single model algorithms (**wRT** - WEKA Regression Tree; **wMT** - WEKA Model Tree; **STRT** - CLUS Single Target Regression Tree; **MTRT** - CLUS Multi target Regression Tree); *b*) Ensemble algorithms (**wBagMT** - WEKA Bag of Model Trees; **BagSTRT** - CLUS Bag of STRTs; **RFSTRT** - CLUS Random Forest of STRTs; **RFMTRT** - CLUS Random Forest of MTRTs)

a) **Single model algorithms**

	Single target			Multi-target
Target	wRT	wMT	STRT	MTRT
H[m]	2.336 ± 0.035	2.271 ± 0.038	2.361 ± 0.025	2.373 ± 0.038
CC[%]	16.068 ± 0.051	15.758 ± 0.129	16.481 ± 0.151	14.708 ± 0.108

b)**Ensemble algorithms**

	Single target			Multi-target	
Target	wBagMT	BagSTRT	RFSTRT	BagMTRT	RFMTRT
H[m]	2.091 ± 0.038	2.071 ± 0.029	2.056 ± 0.030	2.070 ± 0.028	2.054 ± 0.029
CC[%]	14.723 ± 0.079	14.868 ± 0.125	14.713 ± 0.105	14.891 ± 0.109	14.708 ± 0.108

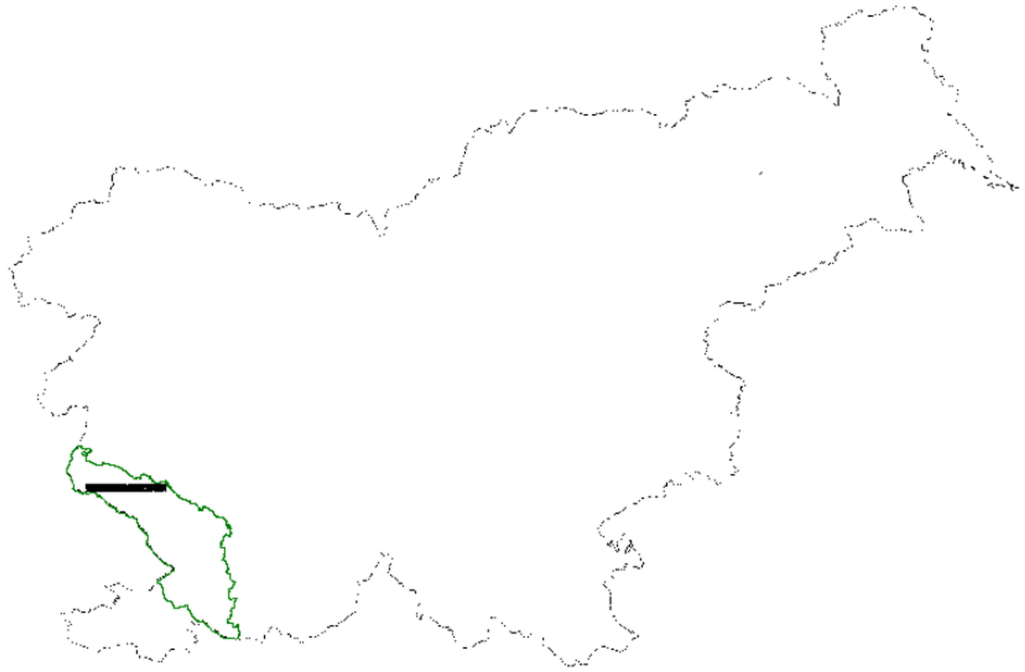


Figure 1: A contour map of Slovenia. The study area is encompassed with a black line whereas the area recorded with LiDAR is presented with black color. The white dots in the LiDAR area present the area not covered with vegetation (e.g., settlements) and these parts were not included in the study.

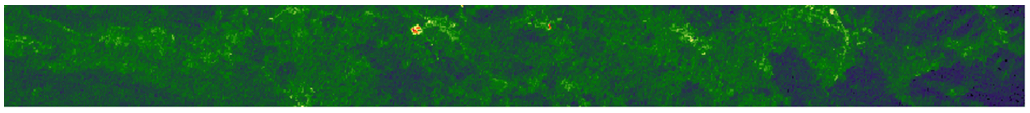
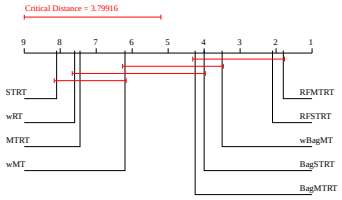
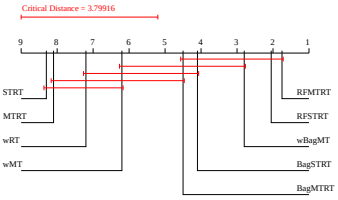


Figure 2: A part of Landsat ETM+ band 3' image that covers the area recorded with LiDAR acquired on 10.11.2002

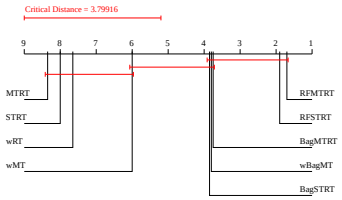
Figure 3: Fine image segmentation of the Landsat ETM+ band 3' image acquired on 10.11.2002 (presented in Figure 2)



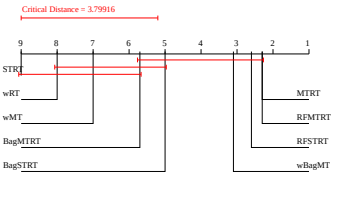
(a) Target variable: **H**; Measure: **Corr**



(b) Target variable: **CC**; Measure: **Corr**



(c) Target variable: **H**; Measure: **RMSE**



(d) Target variable: **CC**; Measure: **RMSE**

Figure 4: Average ranks diagrams: *a*) target variable - H and eval. measure - Corr; *b*) target variable - CC and eval. measure - Corr; *c*) target variable - H and eval. measure - RMSE and *d*) target variable - CC and eval. measure - RMSE. Algorithms with lower ranks (far right) perform better. Algorithms whose average rank difference is larger than the critical difference can be considered significantly different with 95 % probability. The algorithms that do not differ significantly are connected with a line. Algorithm labels are as follows: **wRT** - WEKA Regression Tree; **wMT** - WEKA Model Tree; **STRT** - CLUS Single-target Regression Tree; **MTRT** - CLUS Multi-target Regression Tree; **wBagMT** - WEKA Bag of Model Trees; **BagSTRT** - CLUS Bag of STRTs; **RFSTRT** - CLUS Random Forest of STRTs; **RFMTRT** - CLUS Random forest of MTRTs

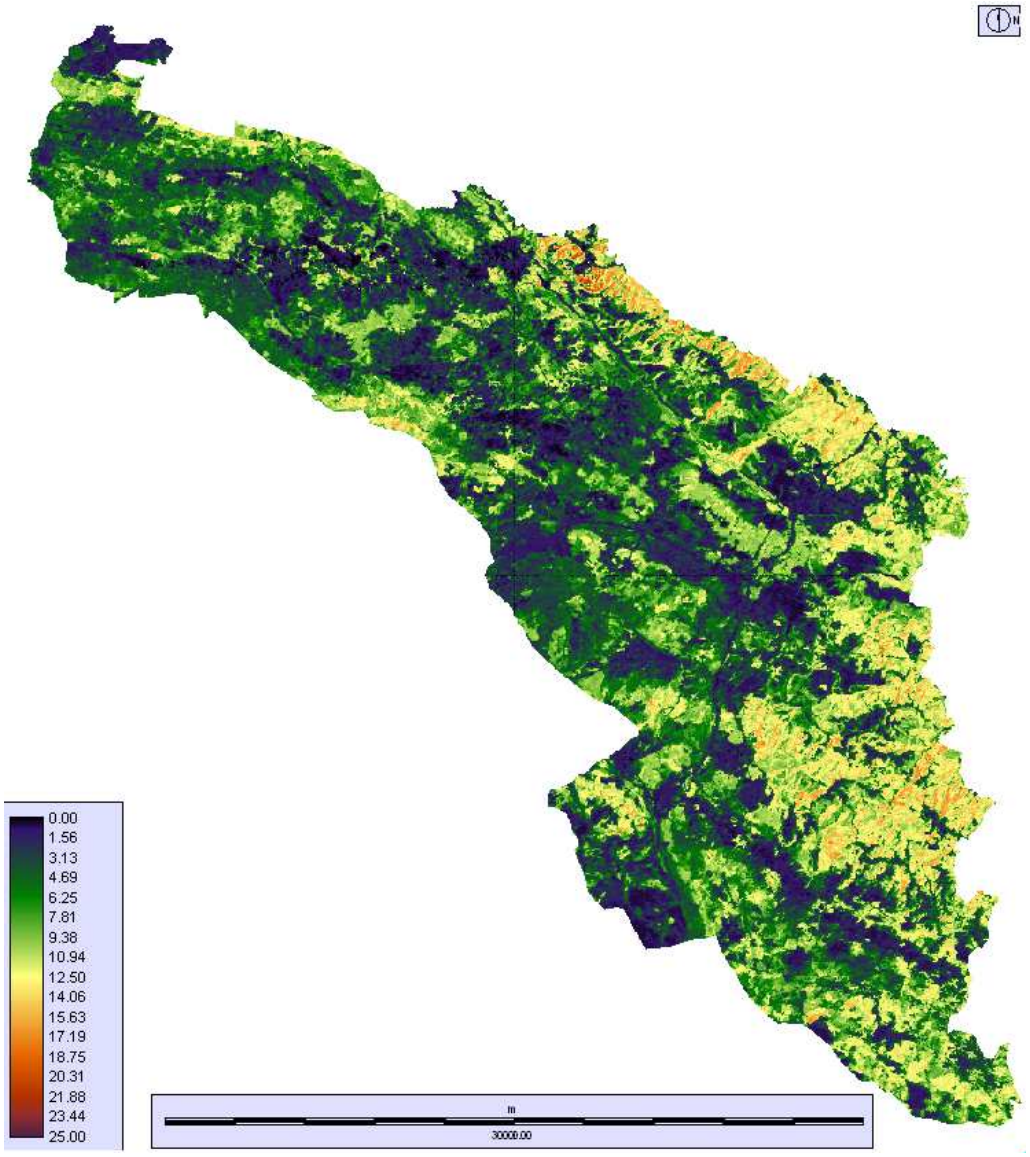


Figure 5: Map of vegetation height for the Kras region generated by using a random forest of multi-target regression trees model. The legend shows the vegetation height in meters.

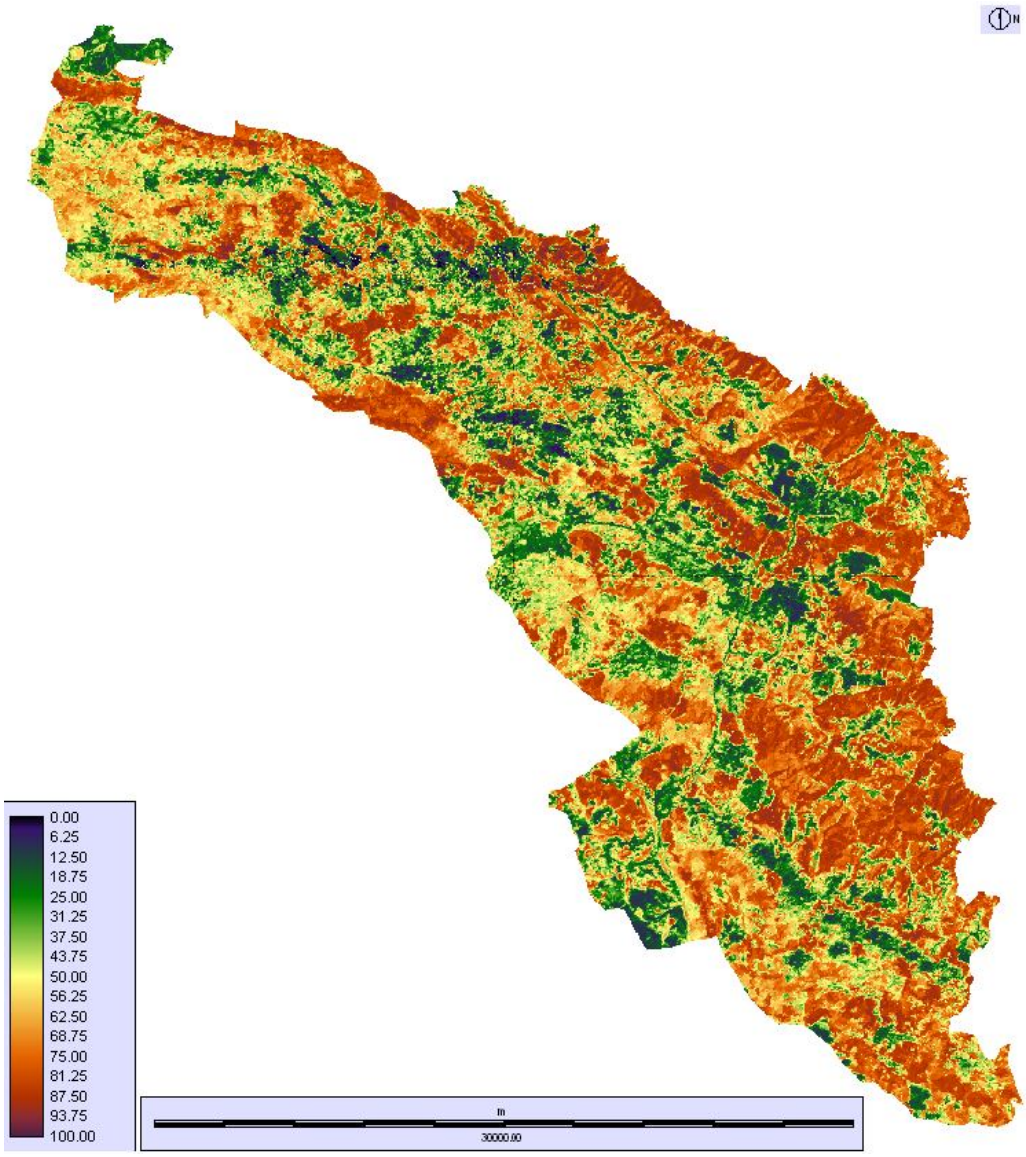


Figure 6: Map of canopy cover for the Kras region generated by using a random forest of multi-target regression trees model. The legend shows the percentage of canopy cover.