Tree Species Classification and Input Data Evaluation

Petra Krahwinkler & Jürgen Rossmann

To cite this article: Petra Krahwinkler & Jürgen Rossmann (2013) Tree Species Classification and Input Data Evaluation, European Journal of Remote Sensing, 46:1, 535-549, DOI: 10.5721/EuJRS20134631

To link to this article: https://doi.org/10.5721/EuJRS20134631

© 2013 The Author(s). Published by Taylor & Francis.

Published online: 17 Feb 2017.

Submit your article to this journal

Article views: 93

View related articles

Citing articles: 1 View citing articles
Tree Species Classification and Input Data Evaluation

Petra Krahwinkler* and Jürgen Rossmann

Institute for Man-Machine Interaction, RWTH Aachen University, Ahornstrasse 55, 52074 - Aachen, Germany
*Corresponding author, e-mail address: petra.krahwinkler@rwth-aachen.de

Abstract
Semantic world models require detailed classifications that allow the generation of entities that represent real world objects. This study combines advanced binary support vector machine classifiers in the hierarchical structure of a binary decision tree. This approach fulfills the requirements of a subsequent semantic model generation approach as it is an object based approach that works on the single tree level. It is fast and can be applied easily without expert knowledge. The algorithm can be trained on existing tree sample inventory data and additional samples from other data sources can be imported or placed. Classification results were calculated on a variety of input data sources and spatial resolutions. The achieved accuracies were analyzed and provide information for the decision support regarding input data choice for future applications.

Keywords: Tree species classification, support vector machines, decision tree, single tree level, very high resolution.

Introduction
Semantic world models describe how the virtual objects relate the real world. Therefore, they are prerequisites for single tree based applications like virtual inventories, forest growth simulations and the use of (semi-)autonomous robots in the complex, unstructured, natural environments like forests. Semantic world models are generated using three abstractions, namely classification, aggregation and generalization as described in Smith and Smith [1977] and ter Bekke [1992]. Spatially explicit databases derived from individual tree species classification are important for a wide variety of forest related activities like forest inventories, diversity monitoring, sustainable forest management and robotic applications in forests. While multiple tree species classification techniques have been described, they typically are performed at resolutions that are too coarse for single tree based analysis [Goodenough et al., 2003; Törmä et al., 2004; Scheuber, 2010]. Other studies used hyperspectral data [Voss and Sugumaran, 2008; Cho et al., 2010], which is too expensive for many applications. Only a small test area was used in Kosaka et al. [2005] and only few tree species were classified in Leckie et al. [2005], Chubey et al. [2006], van Aardt et al.
Commonly used algorithms in these studies include maximum likelihood [Stoffels et al., 2011], spectral angle mapper [Cho et al., 2010], k-nearest neighbor [Tomppo et al., 2008], decision trees [Vikhamar and Fjone, 2004], artificial neural networks [Niska and Packalen, 2010] and support vector machines [Huang et al., 2002; Zortea et al., 2007; Iovan et al., 2008]. Each classifier has its drawback and needs to be chosen with regard to the specific requirements of the application. While k-nearest neighbor classification has been widely used in national forest inventories as described by Tomppo et al. [2008], these applications use low resolution data and focus on estimating an overall distribution of species over a large area. This study on the other hand focuses on single tree level classification and k-nearest neighbor has been shown to perform inferior compared to the proposed support vector machine based decision tree in Krahwinkler et al. [2011]. Decision trees have been used increasingly in a multiple classifier approach called random forest [Immitzer et al., 2012]. It usually incorporates several hundred trees and therefore has to be considered as a black box classifier, which does not provide insight in the nature of the class differences. Other advanced algorithms with good performance in several studies like artificial neural networks and support vector machines also belong to the group of black box classifiers. Another drawback of artificial neural networks is that they do not guarantee to find the optimal solution but support vector machines on the other hand maximize the generalization ability [Vapnik, 2000].

The development of a highly automated classification process with a minimum requirement on user interaction in reference data acquisition, validation and filtering as well as in the classification process itself, are the main objectives of this study. The approach needs to work with multiple very high resolution data sources on large areas to generate classification results on the single tree level. Furthermore, the cost of input data sources has to be kept in mind, which requires an approach that allows the use of different input data sources. With a flexible system that does not require expensive sources like hyperspectral data, input data can be chosen with regard to its benefit-cost ratio.

The classification results from this study will be used to generate a semantic world model as part of the Virtual Forest testbed [Rossmann et al., 2010a], which can be used for robotic applications and forest management activities.

**Study Area and Data Processing**

The test area is located in western Germany near the city of Arnsberg and has a size of approximately 340km² whereof about 230km² are forest areas. An overview of the area is given in Figure 1.

The available data sources are described in Table 1. The normalized digital surface model (nDSM) is calculated as the difference between the digital surface model (DSM) and the digital ground model (DGM).

The nDSM is used to delineate objects for the object based classification using a Gaussian filter and a watershed algorithm. Precise single tree delineation requires a high resolution tree species classification. Object-based approaches have been noted to perform better than pixel-based approaches in Voss and Sugumaran [2008] and Blaschke and Strobl [2001]. In the case of under segmentation, one object might contain two trees, which belong to different species. To avoid this effect, the object delineation approach allows some over segmentation, especially in the case of broad leaved trees. In this case, one tree crown may
be split into several objects, which will then be classified separately.

Figure 1 - Test area Arnsberg, Germany. Geographic location: 51°23'N 8°03'E, size: 340km², north up.

Table 1 - Available data sources. The available bands are the normalized differential surface model (nDSM), red (R), green (G), blue (B), near infrared (NIR), red edge (RE), and short wavelength infrared (SWIR).

| Abbr. | Bands       | Resolution | Sensor                  | Date       |
|-------|-------------|------------|-------------------------|------------|
| nDSM  | nDSM        | 0.4m       | LiDAR Riegl LMS-Q560    | 2008-06    |
| RGB, CIR | R, G, B, NIR | 0.1m       | Vexcel Ultracam-X       | 2008-07-01 |
| SPOT  | NIR, R, G, SWIR | 5m, 10m   | SPOT 5 satellite        | 2008-09-28 |
| REE   | R, G, B, RE, NIR | 5m        | RapidEye satellite      | 2010-06-05 |

The ground reference data contains 211 inventory points, which were measured as part of the federal forest inventory. The points are located in 1km square grid with additional points in 500 m and 250 m square grids. At each of these points, all trees within a 12 m radius and a diameter at breast height larger than 10cm are measured, resulting in a total of 8559 tree samples. Each of these tree samples contains information on the coordinates of the tree, the tree species and the diameter at breast height. Several tree samples also contain height measurements and additional attributes. In addition, 4801 tree samples from eight rather homogeneous fully mapped forest stands, which contain species that are less common in the test area, were available as reference data. The tree sample inventory reference samples needed to be filtered as the field measurements were acquired in 1997 and differs from the
acquisition date of the remote sensing data in 2008 and 2010. The reference data from the fully mapped forest stands also requires some filtering as the airborne visibility was not regarded during field sampling and in some parts errors in the georeferencing occurred. The result of the filtering is shown in Table 2, where OBL is a group of broadleaved species with long rotation time and OBS is a broadleaved species group with short rotation time.

| species | total | good | duplicate | shadow | too low | invalid |
|---------|-------|------|-----------|--------|---------|---------|
| oak     | 520   | 348  | 161       | 3      | 7       | 1       |
| beech   | 2751  | 1134 | 1596      | 4      | 12      | 5       |
| OBL     | 155   | 57   | 98        | 0      | 0       | 0       |
| OBS     | 2516  | 824  | 1673      | 8      | 11      | 0       |
| poplar  | 18    | 6    | 11        | 1      | 0       | 0       |
| pine    | 76    | 32   | 43        | 1      | 0       | 0       |
| larch   | 558   | 292  | 249       | 4      | 10      | 3       |
| spruce  | 6039  | 2128 | 3479      | 134    | 198     | 100     |
| Douglas Fir | 727 | 332  | 384       | 3      | 6       | 2       |

Some of the trees in the sample inventory are not present in the remote sensing data due to felling, depth or storm damage. This results in more than one sample located within an image object, samples within objects dominated by shadows or samples where the measured tree height does not correspond to the height in the nDSM. Filtering for duplicates and too low samples will lead to a better representation of the tree species in the test area and not decrease classification accuracy. Duplicates and low samples are not visible in the remote sensing data. They represent data that cannot be seen from above as they belong to understory trees and are not part of the research objective of this study. Shadow samples do pose a problem for classification accuracy. In these areas the reflectance is low and therefore the information contained in the remote sensing data on these areas is limited. Some studies have introduced shadow classes into their classification scheme. We only filtered very dark areas and kept the samples within underexposed areas with their original classification. But we did find that classification errors increased in underexposed areas as several objects were misclassified as spruce. The inventory data does not contain non-tree samples. An additional class of non-tree samples has been introduced and almost 100% accuracy in separating forest and non-tree samples was reached. However, the focus of this study is to study the classification of tree species and therefore non-forest samples, were not included in this study.

In a ten-fold cross validation approach, the remaining tree samples are grouped ten times into pairs of test and training sets such that each pair is mutually exclusive but samples may be used in several of the ten pairs. Only species with at least 150 samples are used, whereof at least 50 samples are spared for testing as proposed by Congalton and Green [2009]. Of the nine species groups used in stand-based forest inventory in North Rhine-Westphalia, six are represented by a sufficient number of samples, namely oak, beech, OBS, larch, spruce
and Douglas fir. These are the species that were classified in this study. The classification scheme used in this study is therefore incomplete which means that the final map accuracy will most likely be lower than the accuracies presented here. For practical application the underrepresented species should be assigned to one of the remaining tree species to create a complete classification scheme and to be able to estimate the final map accuracy.

A well-known problem in multi sensor based studies is the existence of co-registration errors. The data sources used here had very high precision. Each data source was inspected individually regarding its co-registration and necessary corrections were performed. The features for each object were extracted for each data source individually to calculate the object feature vector that is used in the classifier training and classification methods. For the tree samples, co-registration was addressed by the filtering approach.

**Support Vector Machine based Decision Tree**

Burges [1998] and Vapnik [2000] described support vector machines. They are a very powerful binary, supervised classification. The generalization error is minimized by maximizing the margin along the decision boundary between the two classes by solving the convex quadratic programming problem as given in [Burges, 1998].

\[
W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \]

with \(0 \leq \alpha_i \leq C; i = 1..l\) \[1\]

and \( \sum_{i=1}^{l} \alpha_i y_i = 0 \)

The Lagrange multipliers are denoted by \(\alpha_i, y_i \in [-1,1]\) are the class assignments of the samples and \(x_i\) are the feature vectors in input space. \(K(x_i \cdot x_j) = \|x_i - x_j\|^2\) is the kernel function, in this case a radial basis function kernel. \(C\) is the cost parameter corresponding to the penalty for errors in the case of non-separable data, while \(\gamma\) corresponds to the width of the radial basis function. \(C\) and \(\gamma\) are parameters estimated during training using grid search in the range of \(C = 2^{-5}, 2^{-3}, \ldots, 2^{13}\) and \(\gamma = 2^{-15}, 2^{-13}, \ldots, 2^3\) and 10-fold cross validation. In general, small values of \(C\) and \(\gamma\) are preferred as they avoid overtraining of the support vector machine. Several multiclass extensions exist. The most common are the one-against-one and the one-against-all approaches. In this study, a support vector machine based binary decision tree was used. Similar approaches have been used in Takahashi and Abe [2002], Chen et al. [2008], Dong and Chen [2008], Madzarov and Gjorgjevikj [2009]. The support vector machine based decision tree used here has a hierarchical structure, which gives information about the spectral similarity of species and allows collapsing or expanding tree nodes without having to retrain the whole tree. At each node, all possible combinations of separating the classes into two groups of almost equal size are determined. For each of these pairs of sets, the equivalent distance is
calculated as described in Dong and Chen [2008]:

\[ d_E = \sqrt{\sum_{i=0}^{N} \left( \frac{\mu_{1,i} - \mu_{2,i}}{\sigma_{1,i} + \sigma_{2,i}} \right)^2} \]  \[ [2] \]

The pair with the highest equivalent distance is chosen for the separation at the current node. This approach guarantees a balanced tree which for \( n = 9 \) classes has \( N = 2n - 1 = 17 \) tree nodes, whereof \( n - 1 \) nodes are inner tree nodes. For each of these inner tree nodes a support vector machine is trained. At each inner tree node the 10-fold cross validation is calculated and gives information on the separability of the species groups. An example of a support vector machine based decision tree is given in Figure 2 with the cross-validation results given for each inner tree node. The number of support vector machines that need to be evaluated for each classification is \( \lceil \log_2(N) \rceil = 4 \).

![Figure 2 - Example of a support vector machine based decision tree. Each inner tree node contains a binary support vector machine. The 10-fold cross validation results for each SVM are given for each inner tree node.](image)

The support vector machine based decision tree is trained on the input data sources described above. The approach has been proven to outperform k-nearest neighbor classification [Krahwinkler et al., 2011] and has also shown to yield slightly better results than a one-against-one multiclass support vector machine approach while reducing classification time by a factor of about 0.7. Furthermore, the support vector machine based decision tree offers information on the hierarchy in the spectral similarity of the used classes. It can be expanded and collapsed to some extent without retraining the whole tree.

**Comparisons and Discussion**

The quality and availability of input data sources and reference data are vital for the achieved classification accuracies. New developments lead to new data sources with higher resolutions and better data availability. Nevertheless, higher resolutions for example still come with a higher cost and it is important to evaluate the value of the increased resolution. Therefore, the accuracies
achieved on several resolutions, input data sources and numbers of training data were compared. For the comparison of input data resolutions, only LiDAR nDSM and airborne RGB and CIR images were used. Only these data sources were available at a resolution 0.4 m and 0.1 m respectively and therefore allow the comparison of the classification accuracies achieved on these high resolutions. The RapidEye and SPOT satellite data have resolutions of only 5 m and 10 m respectively. The airborne images have an original resolution of 0.1 m per pixel and were resampled to several lower resolutions. The LiDAR nDSM data have a resolution of 0.4 m per pixel. The training sets contained 182 larches, 792 spruces, 182 Douglas firs, 250 oaks, 548 beeches and 182 other broadleaved samples. The test set included 41 larches, 608 spruces, 19 Douglas firs, 83 oaks, 272 beeches and 51 other broadleaved samples. In this paper, it was not possible to give all calculated confusion matrices as proposed by Congalton and Green [2009]. Therefore, the classification accuracies were estimated using ten-fold cross validation on ten stratified pairs of training and test sets extracted from one data set, following the approach described in Witten and Frank [2005]. From the individual results, the mean and standard deviation of the overall accuracy as well as the mean and standard deviation of the species-wise producer’s and the user’s accuracies were calculated. Although the standard deviation does not give a confidence interval for the calculated mean cross validation results, it gives an estimate of the variation of the ten cross validation calculations. The Kappa coefficient has been criticized in Foody [2011] and Pontius and Millones [2011]. As it is still used in many studies, we also calculated the Kappa values. The results are given in Table 3, where the standard deviation is given in percentage points (pp).

Table 3 - Comparison of classification accuracies calculated for ten pairs of test and training sets, subject to input data resolutions. Std. is the standard deviation and prod. denotes the producer’s accuracy.

|               | 50 m   | 20 m   | 10 m   | 5 m    | 2 m    | 1 m    | 0.4 m  | 0.2 m  |
|---------------|--------|--------|--------|--------|--------|--------|--------|--------|
| mean          | 68.9%  | 70.2%  | 73.6%  | 73.7%  | 73.0%  | 72.9%  | 72.8%  | 72.7%  |
| std.          | 1.8pp  | 1.7pp  | 1.7pp  | 1.0pp  | 1.3pp  | 0.8pp  | 1.2pp  | 1.2pp  |
| prod. mean    | 51.9%  | 56.8%  | 60.1%  | 56.9%  | 57.1%  | 56.5%  | 55.6%  | 54.8%  |
| prod. std.    | 1.9pp  | 2.1pp  | 2.6pp  | 2.3pp  | 2.2pp  | 1.3pp  | 1.8pp  | 1.6p   |
| user mean     | 52.7%  | 51.3%  | 55.8%  | 53.6%  | 52.3%  | 51.5%  | 50.9%  | 50.7%  |
| user std.     | 3.9pp  | 3.3pp  | 3.0pp  | 2.4pp  | 2.2pp  | 1.5pp  | 1.7pp  | 2.0pp  |
| Kappa mean    | 0.489  | 0.525  | 0.577  | 0.575  | 0.565  | 0.563  | 0.560  | 0.558  |
| Kappa std.    | 0.025  | 0.022  | 0.025  | 0.013  | 0.018  | 0.011  | 0.017  | 0.017  |

Table 3 shows that very high resolutions of less than 1 m do not lead to improved classification results in this test case. One reason is that very high resolution data leads to high within-crown variability. The differences between the classification results at a 5 m and a 0.1 m resolution in the overall accuracy are within the range of the standard deviation, but the mean user’s accuracy drops from 53.6% to 50.7% with increasing resolution. Higher resolution also leads to increased classification times as the data access time becomes a critical factor. Increasing the resolution from 0.4 m to 0.2 m increased the classification time for the given test area by a factor of about 1.27 and the increase from 0.2 m to 0.1 m lead to an approximately 5 times increased calculation time in our setup with a standard PC and the raster sources stored in the intranet with a 1 GB local area network connection in between.

At a resolution below 20 m, the accuracy drops significantly. At this resolution, single tree based classification is not possible anymore as single crowns are not visible in the data and
spectral mixing within pixels increases. Homogeneous forest stands can still be classified correctly, but some trees within highly mixes forest stands will be misclassified. The confusion matrix for one test case at a resolution of 5 m is given in Table 4.

Table 4 - Confusion matrix at an image resolution of 5m per pixel. OBS describes a group of other broadleaved species with short rotation time.

|       | oak | beech | OBS | larch | spruce | Douglas fir | user’s |
|-------|-----|-------|-----|-------|--------|--------------|--------|
| oak   | 16  | 18    | 1   | 0     | 10     | 0            | 35.6%  |
| beech | 47  | 205   | 12  | 5     | 47     | 0            | 64.9%  |
| OBS   | 5   | 15    | 31  | 1     | 7      | 0            | 52.5%  |
| larch | 3   | 7     | 1   | 20    | 17     | 0            | 41.7%  |
| spruce| 11  | 22    | 6   | 14    | 491    | 4            | 89.6%  |
| Douglas fir | 1 | 5 | 0 | 1 | 36 | 15 | 25.9% |
| producer’s | 19.3% | 75.4% | 60.8% | 48.8% | 80.8% | 78.9% | 72.4% |

Although the overall accuracy of 72.4% is rather high, the species-wise user’s and producer’s accuracies are rather low with a mean producer’s accuracy of 60.7% and a mean user’s accuracy of 51.7%.

To increase the overall accuracy as well as the species-wise user’s and producer’s accuracies, additional data sources are required as input data sources. Airborne LiDAR data as well as RGB and CIR images are often acquired by national agencies or land surveying offices and in some countries can be obtained for free by national agencies and research institutions. However, the spectral and temporal resolution of airborne RGB and CIR images is limited to 4 spectral bands and one acquisition date. Therefore, additional satellite data can improve classification results due to additional spectral bands and additional acquisition dates, which might be more suitable for tree species classification and allow multitemporal analysis. The increased availability of satellite sensor systems now offers a wide range of data with different sensor specifications and at relatively low costs. The choice of the sensor and the phenological conditions at acquisition time will highly influence the achieved classification results. Satellite sensor data can also provide increased homogeneity compared to airborne data as the acquisition of larger areas within one day is possible, thereby avoiding different atmospheric conditions within the area of interest. Table 5 shows a comparison of classification results based on airborne images, two types of satellite sensor data, and combinations of these input data sources.

Table 5 - Comparison of classification accuracies subject to input data sources excluding airborne LiDAR data.

|       | RGB +CIR | REE | SPOT | RGB +CIR +REE | RGB +CIR +SPOT | RGB +CIR +REE +SPOT |
|-------|-----------|-----|------|---------------|----------------|----------------------|
| mean  | 71.0%     | 78.0% | 70.3% | 81.0%         | 75.4%          | 82.7%                |
| std.  | 1.1pp     | 1.2pp | 1.4pp | 1.3pp         | 1.4pp          | 1.3pp                |
| prod. mean | 51.4% | 63.0% | 42.2% | 73.5%         | 63.4%          | 76.5%                |
| prod. std. | 3.3pp | 2.0pp | 2.3pp | 3.0pp         | 1.8pp          | 2.8p                 |
| user mean | 46.6% | 56.5% | 41.0% | 63.9%         | 55.3%          | 67.6%                |
| user std. | 4.6pp | 2.4pp | 4.1pp | 2.7pp         | 2.1pp          | 2.9pp                |
| Kappa mean | 0.529 | 0.660 | 0.506 | 0.699         | 0.611          | 0.725                |
| Kappa std. | 0.015 | 0.019 | 0.016 | 0.020         | 0.020          | 0.019                |

The highest classification accuracy is reached for the combination of all available data sources in the last column of Table 5. The addition of the RapidEye data set to the airborne image data.
improves the overall accuracy by 10 percentage points, while the SPOT data set only leads to an improvement by 4.4 percentage points. The SPOT satellite data improves the classification based on the airborne images and RapidEye data sets by another 1.7 percentage points. The lowest accuracy is reached on the SPOT data set alone, which only has a resolution of 10 m in the red, green and near infrared bands and only 20 m resolution in the short wavelength infrared band.

Airborne LiDAR nDSM is a rather expensive data source. However, the tree height can be used as a measure describing the age of the trees. As the spectral response of tree crowns changes with age, the additional height feature provided by the nDSM can help in the discrimination of tree species. Table 6 shows the results of the classifications that were reached by using the LiDAR nDSM data set in addition to the airborne and satellite data sources. Although adding the LiDAR nDSM data increases the overall accuracy clearly for the calculations based on single spectral input data, the increase is less apparent for the combinations. For the calculations including the airborne and RapidEye data sets, the increase due to the nDSM data is within the range of the standard deviation.

Table 6 - Comparison of classification accuracies subject to airborne LiDAR data and additional input data sources.

| nDSM     | +RGB+CIR | +RE  | +SPOT | +RGB +CIR |
|----------|----------|------|-------|-----------|
|          |          |      |       |           |
| mean     | 73.7%    | 79.9%| 72.3% | 81.7%     |
| std.     | 1.1pp    | 1.2pp| 1.0pp | 1.2pp     |
| prod. mean | 56.9% | 68.6%| 47.7% | 73.8%     |
| prod. std.| 2.6pp    | 3.1pp| 3.0pp | 2.2pp     |
| user mean | 53.6%    | 60.1%| 45.6% | 64.9%     |
| user std. | 2.4pp    | 2.5pp| 2.6pp | 2.1pp     |
| Kappa mean | 0.576  | 0.543| 0.699 | 0.705     |
| Kappa std.| 0.014    | 0.014| 0.020 | 0.006     |

The detailed results for one test set using LiDAR nDSM, airborne RGB and CIR images and SPOT and RapidEye satellite data, the combination with the highest classification accuracy, is shown in Table 7.

Compared to Table 4 it can be noted, that not only the overall classification accuracy increased but also all the species-specific user’s and producer’s accuracies increased significantly. The user’s accuracies for the individual classes increased by values between 3.8 pp for larch and 39.5 pp for Douglas fir and the producer’s accuracies increased in the range of 5.9 pp for OBS and 65.1 pp for oak. The overall accuracy increases by 12.3 pp from 72.4% to 84.7% due to the addition of the SPOT and RapidEye satellite data sets.

Table 7 - Confusion matrix using LiDAR nDSM, airborne RGB and CIR, and SPOT and RapidEye satellite data.

|         | oak | beech | OBS | larch | spruce | Douglas fir | user's |
|---------|-----|-------|-----|-------|--------|-------------|--------|
| oak     | 70  | 17    | 3   | 0     | 5      | 0           | 73.7%  |
| beech   | 5   | 230   | 6   | 2     | 35     | 0           | 82.7%  |
| OBS     | 3   | 8     | 34  | 1     | 9      | 0           | 61.8%  |
| larch   | 4   | 9     | 2   | 30    | 21     | 0           | 45.5%  |
| spruce  | 1   | 8     | 6   | 8     | 529    | 2           | 95.5%  |
| Douglas fir | 0   | 0     | 0   | 9     | 17     | 65.4%       |
| producer's | 84.3% | 84.6% | 66.7% | 73.2% | 87.0% | 89.5% | 84.7% |
However, it is unclear to which parts the increase in classification accuracies is caused by the multitemporal information, the additional spectral bands, better suitability of the acquisition dates, or better homogeneity of the RapidEye and SPOT data sets. Another very important factor for practical applications is the minimum number of reference samples that are required for classifier training and accuracy assessment. Congalton [1999] suggests a minimum number of 50 samples per species for testing if the test area is less than one million acres and the number of classes is less than 12. Both conditions are fulfilled in this test case. Witten and Frank [2005] state, that in practical terms it is common to hold out one-third of the data for testing and use the remaining two-thirds for training. Combining these rules leads to a minimum number of 150 samples per species. For six species, this leads to an overall number of 900 samples.

The acquisition of accurate reference data is costly and time-consuming. Therefore, reference data requirements should be as low as possible for practical applications. To estimate the effect of limited reference data, ten training and test sets were extracted. To create sets with reduced numbers of training data, only a fraction of the original training data was used while the test samples stayed exactly the same. The original training sets with at least 182 samples per species contained 182 larches, 792 spruces, 182 Douglas firs, 250 oaks, 548 beeches and 182 other broadleaved species. The test set contains 41 larches, 608 spruces, 19 Douglas firs, 83 oaks, 272 beeches and 51 other broadleaved species, thereby weighting the species according to their presence in the test area. The minimum number of 50 test samples is not fulfilled for larch and Douglas fir in this test set configuration as the sample numbers are adjusted to the actual species distribution in the test area to represent an estimate of the resulting map accuracy. For all results in Table 8, RapidEye and SPOT satellite data were used in addition to the airborne images and LiDAR data.

| number | 72 | 84 | 109 | 133 | 158 | 182 |
|---|---|---|---|---|---|---|
| mean | 71.5% | 75.0% | 80.8% | 82.2% | 82.8% | 83.8% |
| std. | 2.4pp | 1.9pp | 1.6pp | 1.1pp | 1.4pp | 1.0pp |
| prod. mean | 73.5% | 73.6% | 74.9% | 76.0% | 76.5% | 77.9% |
| prod. std. | 2.6pp | 2.9pp | 2.3pp | 3.1pp | 3.0pp | 2.1pp |
| user mean | 53.8% | 55.9% | 63.4% | 65.4% | 66.7% | 67.6% |
| user std. | 1.9pp | 2.1pp | 2.7pp | 2.0pp | 2.3pp | 1.9pp |

Figure 3 shows the accuracy values for different sample sizes as given in Table 8. The standard deviation is shown as error indicators.

In Table 8 and Figure 3 it can be seen that the overall accuracy decreases slightly by 3 pp from 182 to 109 samples and declines more rapidly by 9.3 pp from 109 to 72 samples. A similar dip can be noticed in the user’s accuracy, which decreases by 4.2 pp from 182 to 109 samples and drops by 9.6 pp from 109 to 72 samples. The producer’s accuracy decreases more uniformly in the whole range by 4.4 pp from 182
to 72 samples. These results lead to the conclusion that at least 109 samples should be used for training to avoid the drop in classification accuracy but higher numbers are preferably. The results support the assumptions of using at least 150 samples per species, whereof 50 are spared for testing and at least 100 are used for classifier training.

![Figure 3 - Accuracies for different training sample sizes. The standard deviation is given as error indicators.](image)

**Conclusions**

The proposed approach requires reference data that contain information about the coordinate and the tree species. Existing tree sample inventory data can be used and may be supplemented by additional reference data points and visually placed samples. However, the quality of the reference data influences the result substantially. The ideal data source would be random distribution of the samples in the whole test area but such a data source is seldom available and difficult and expensive to acquire. This has to be kept in mind when choosing reference data sources and planning field measurements. Furthermore, it requires spectral input data. Several input data sources including LiDAR nDSM, LiDAR intensity data, airborne image data and satellite images can be combined. The comparisons of input data resolutions, input data sources and number of reference data samples offer important information for decision support and design of future single tree level based tree species classification applications.

The approach has been proven to work well in creating classification results as a basis for semantic world modeling as described in Rossmann et al. [2010b] and Rossmann et al. [2011b]. Numerous applications can be implemented using this semantic world model, including single tree based forest inventory, forest growth simulation, navigation and harvesting simulation [Rossmann et al., 2010c]. Figure 4 shows an example of a wood harvester simulation. Two laser scanners and the tree species classification based single tree level semantic world model are used for navigation as GPS suffers from low position accuracy under closed canopy as described in Rossmann et al. [2011a].
Figure 4 - Wood harvester simulation and laser scanner based navigation using a semantic world model based on the generated single tree level species classification.

Acknowledgements
This project is co-financed by the European Union and the federal state of North Rhine-Westphalia, European Regional Development Fund (ERDF). Europe - Investing in our future.

References
Blaschke T., Strobl J. (2001) - What’s wrong with pixels? Some recent developments interfacing remote sensing and GIS. GIS: Zeitschrift für Geoinformationssysteme, 14 (6): 12-17. ISSN: 0935-1523.
Burges C. (1998) - A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery, 2 (2): 121-167. doi: http://dx.doi.org/10.1023/A:1009715923555.
Chen J., Wang C., Wang R. (2008) - Combining Support Vector Machines With a Pairwise Decision Tree. IEEE Geoscience and Remote Sensing Letters, 5 (3): 409-413. doi: http://dx.doi.org/10.1109/LGRS.2008.916834.
Cho M.A., Debba P., Mathieu R., Naidoo L., van Aardt J.A., Asner G.P. (2010) - Improving Discrimination of Savanna Tree Species Through a Multiple-Endmember Spectral Angle Mapper Approach: Canopy-Level Analysis. IEEE Transactions on Geoscience and Remote Sensing, 48: 4133-4142. doi: http://dx.doi.org/10.1109/TGRS.2010.2058579.
Chubey M.S., Franklin S.E., Wulder M.A. (2006) - Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. Photogrammetric Engineering and Remote Sensing, 72 (4): 383-394. ISSN: 0099-1112.
Congalton R.G. (1999) - A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment, 37 (1): 35-46. doi: http://dx.doi.org/10.1016/0034-4257(91)90048-B.
Congalton R.G., Green K. (2009) - Assessing the Accuracy of Remotely Sensed Data, 2nd ed. Taylor & Francis, p. 183.

Dong G., Chen J. (2008) - Study on Support Vector Machine Based Decision Tree and Application. Proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 5: 318-322. doi: http://dx.doi.org/10.1109/FSKD.2008.252.

Foody G.M. (2011) - Classification Accuracy Assessment. IEEE Geoscience and Remote Sensing Society Newsletter, 159: 8-14. ISSN: 0274-6338.

Goodenough D.G., Dyk A., Niemann K.O., Pearlman J.S., Chen H., Han T., Murdoch M., West C. (2003) - Processing Hyperion and ALI for forest classification. IEEE Transactions on Geoscience and Remote Sensing, 41 (6): 1321-1331. ISSN 0196-2892.

Huang C., Davis L., Townshend J. (2002) - An assessment of support vector machines for land cover classification. International Journal of Remote Sensing, 23 (4): 725-749. doi: http://dx.doi.org/10.1080/01431160110040323.

Immitzer M., Atzberger C., Koukal T. (2012) - Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. Remote Sensing, 4 (9): 2661-2693. doi: http://dx.doi.org/10.3390/rs4092661.

Iovan C., Boldo D., Cord M. (2008) - Detection, Characterization, and Modeling Vegetation in Urban Areas From High-Resolution Aerial Imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1: 206-213. doi: http://dx.doi.org/10.1109/JSTARS.2008.2007514.

Kosaka N., Akiyama T., Tsai B., Koyima T. (2005) - Forest Type Classification Using Data Fusion of Multispectral and Panchromatic High-Resolution Satellite Imagery. IEEE International Geoscience and Remote Sensing Symposium, 4: 2980-2983. doi: http://dx.doi.org/10.1109/IGARSS.2005.1525695.

Krahwinkler P.M., Rossmann J., Sondermann B. (2011) - Support vector machine based decision tree for very high resolution multispectral forest mapping. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium IGARSS 2011, pp. 43-46. doi: http://dx.doi.org/10.1109/IGARSS.2011.6048893.

Leckie D.G., Gougeon F.A., Tinis S., Nelson T., Burnett C. N., Paradine D. (2005) - Automated tree recognition in old growth conifer stands with high resolution digital imagery. Remote Sensing of Environment, 94 (3): 311-326. doi: http://dx.doi.org/10.1016/j.rse.2004.10.011.

Madzarov G., Gjorgjevikj D. (2009) - Multi-class classification using support vector machines in decision tree architecture. IEEE EUROCON 2009, pp. 288-295. doi: http://dx.doi.org/10.1109/EURCON.2009.5167645.

Niska H., Skon J.-P., Packalen P. (2010) - Neural Networks for the Prediction of Species-Specific Plot Volumes Using Airborne Laser Scanning and Aerial Photographs. IEEE Transactions on Geoscience and Remote Sensing, 48 (3): 1076-1085. doi: http://dx.doi.org/10.1109/TGRS.2009.2029864.

Pontius R.G., Millones M. (2011) - Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing, 32 (15): 4407-4429. doi: http://dx.doi.org/10.1080/01431161.2011.552923.

Rossmann J., Schluse M., Hoppen M., Waspe R. (2010a) - GIS-Based Virtual Testbeds and their Application to Forestry and City Simulation. Proceedings of the ASME 2010 World Conference on Innovative Virtual Reality (WINVR 2010), pp. 61-68. doi: http://
Rossmann J., Schluse M., Schlette C. (2010b) - *The virtual forest: Robotics and simulation technology as the basis for new approaches to the biological and the technical production in the forest*. Journal of Systemics, Cybernetics & Informatics, 8 (5): 43-48. ISSN: 1690-4532.

Rossmann J., Jung T.J., Rast M. (2010c) - *Developing Virtual Testbeds for mobile robotic applications in the woods and on the moon*. 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4952-4957. doi: http://dx.doi.org/10.1109/IROS.2010.5651665.

Rossmann J., Schlette C., Emde M., Sondermann B. (2011a) - *Advanced Self-Localization and Navigation for Mobile Robots in Extraterrestrial Environments*. Computer Technology and Application, 2 (5): 344-353. ISSN 1934-7332.

Rossmann J., Schluse M., Waspe R., Mooshammer R. (2011b) - *Simulation in the Woods: From Remote Sensing based Data Acquisition and Processing to Various Simulation Applications*. Proceedings of the 2011 Winter Simulation Conference, pp. 984-996. doi: http://dx.doi.org/10.1109/WSC.2011.6147823.

Scheuber M. (2010) - *Potentials and limits of the k-nearest-neighbour method for regionalising sample-based data in forestry*. European Journal of Forest Research, 129 (5): 825-832. ISSN 1612-4669.

Smith J.M., Smith D.C.P. (1977) - *Database abstractions: aggregation and generalization*. ACM Transactions on Database Systems, 2 (2): 105-133. doi: http://dx.doi.org/10.1145/320544.320546.

Stoffels J., Mader S., Hill J., Werner W., Ontrup G. (2011) - *Satellite-based stand-wise forest cover type mapping using a spatially adaptive classification approach*. European Journal of Forest Research, 131 (4): 1-19. doi: http://dx.doi.org/10.1007/s10342-011-0577-2.

Takahashi F., Abe S. (2002) - *Decision-tree-based multiclass support vector machines*. Proceedings of the 9th International Conference on Neural Information Processing, ICONIP’02, 3: 1418-1422. doi: http://dx.doi.org/10.1109/ICONIP.2002.1202854

Törmä M., Lumme J., Pyysalo U., Patrikainen N., Luojus K. (2004) - *Tree Species Classification Using ERS SAR and MODIS NDVI Images*. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, 35 (7): 927-932. ISSN 1682-1750.

ter Bekke J.H. (1992) - *Semantic Data Modeling*. Prentice Hall International, pp. 350.

Tomppo E., Olsson H., Staahl G., Nilsson M., Hagner O., Katila M. (2008) - *Combining national forest inventory field plots and remote sensing data for forest databases*. Remote Sensing of Environment, 112 (5): 1982-1999. doi: http://dx.doi.org/10.1016/j.rse.2007.03.032.

van Aardt J.A., Wynne R.H., Scrivani J.A. (2008) - *Lidar-based Mapping of Forest Volume and Biomass by Taxonomic Group Using Structurally Homogenous Segments*. Photogrammetric Engineering and Remote Sensing, 74 (8): 1033-1044. ISSN: 0099-1112.

Vapnik V.N. (2000) - *The nature of statistical learning theory*. Statistics for Engineering and Information Science. New York: Springer, 2nd ed. doi: http://dx.doi.org/10.1007/978-1-4757-3264-1.

Vikhamar D., Fjone G. (2004) - *Mapping forest parameters using decision trees and
multitemporal Landsat ETM+ data. Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2004, pp. 2365-2368. doi: http://dx.doi.org/10.1109/IGARSS.2004.1369763.

Voss M., Sugumaran R. (2008) - The Seasonal Effect on Tree Species Classification in an Urban Environment Using Hyperspectral Data, LiDAR, and an Object-Oriented Approach. Sensors, 8 (5): 3020-3036. doi: http://dx.doi.org/10.3390/s8053020.

Witten I.H., Frank E. (2005) - Data mining: practical machine learning tools and techniques. San Diego, Academic Press.

Zortea M., De Martino M., Serpico S. (2007) - A SVM ensemble approach for spectral-contextual classification of optical high spatial resolution imagery. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2007, pp. 1489-1492. doi: http://dx.doi.org/10.1109/IGARSS.2007.4423090.

© 2013 by the authors; licensee Italian Society of Remote Sensing (AIT). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).