ANALYSIS OF HYPERSPECTRAL AND HIGH-RESOLUTION DATA FOR TREE SPECIES CLASSIFICATION

Petra Krahwinkler, Research Assistant
Jürgen Rossmann, Director
Institute of Man-Machine Interaction, RWTH Aachen University
Ahornstrasse 55
52074 Aachen, Germany
krahwinkler@mmi.rwth-aachen.de
rossmann@mmi.rwth-aachen.de

ABSTRACT

Current tree species classification algorithms often use high-resolution satellite data and are in many cases based on forest stands. The spectral bands of the sensors used for data acquisition are given and cannot be chosen regarding the needs of tree species classification. Furthermore distinction is often limited to deciduous trees, coniferous trees and other land use classes.

Single tree based tree species recognition needs very high resolution data. Classification of several distinct species based only on spectral data requires an analysis of the available data to extract information on the spectral signatures of tree species and to find the bands which contain the most relevant information. Only limited information regarding the signatures of tree species in airborne images can be found in the literature, especially regarding those sorts common in central Europe.

In this paper, 235 bands in the near infrared and short wavelength infrared regions are evaluated and compared to RGB and color infrared images (CIR) as well as SPOT4 images. The spectral signatures and the overlaps in the spectral signatures of different tree species are analyzed.

Using this information, suitable algorithms can be developed and existing algorithms can be evaluated. Furthermore the decision making process for a specific data source can be supported and the gained information can be used in the developments of new sensors to fulfill the requirements of tree species classification.

INTRODUCTION

Sustainable forest management is an important strategy in both preserving biodiversity in forest areas and environmentally compatible forestry to provide wood for industry. Especially in habitat monitoring, detailed information on the forest and the existing tree species is needed. Field measurements are both expensive and time consuming and the area that can be managed is strictly limited to a manageable size depending on the number of staff members who work on obtaining the field data. Remote sensing offers affordable data sets on large areas. Although research on tree species classification based on remote sensing as well as airborne image and laser scanner data has been conducted and several approaches have been tested, various problems remain unsolved. Ground truth calibration and reference data needs to be available on a large area in order to train and evaluate the developed classification algorithms. Furthermore, the algorithms are based on very different data sets which make it hard to compare different approaches. Based on a data set of RGB and color-infrared (CIR) images tree species classification is limited to a small number of tree species. Additional data, like hyperspectral sensors or multitemporal imaging, can be used for classification. (Key, 2001) found that multitemporal images are less significant in improving classification results than additional spectral information.

(Pinard, 2003) analyzed the discrimination of 6 tree species and herbaceous vegetation. However, spectral variability within the same species was not taken into consideration. Instead only one sample per tree species was used for the analysis. Those samples were acquired from measurements with a hand-held spectroradiometer in the crown where 20 measurements were taken and the average was calculated.

(van Aardt, 2000) performed measurements with a spectroradiometer from directly above the crown where one reading consisted of the average of 10 samples. Several data points were collected per tree species. The variability within tree species was taken into account and a stepwise discriminant analysis was performed.

(Gong 1997) state that a smaller number of spectral bands can generate more accurate identification than the use of all available hyperspectral bands of a given data set. Their experiments indicate that the visible bands yield more information for discrimination in forest areas than near-infrared bands.

(Yu 1999) state that the linear discriminant analysis fails when faced with high dimensions of the feature vector and high correlation, as it is the case with adjacent spectral bands. They use a penalized discriminant analysis described in (Hastie, 1995).

This analysis is part of the project Virtual Forest*, which aims at providing a database containing stand-based as well as single tree based information that can be used for forest planning, management and monitoring. We compared different data sources and spectral regions in order to gain a better understanding of the properties and differences of the spectra of several tree species and the differences between data sources. To identify relevant spectral bands is an important prerequisite for the decision making process in finding a suitable sensor respectively a suitable combination of available remote sensing data at a preferable low cost for large area tree species classification.

STUDY SITE, SENSORS AND REFERENCE DATA

The study site is located north-east of the city of Arnsberg in Germany. Figure 1 shows an overview of CIR images in that area and the green polygon shows the area where the hyperspectral data set is available.

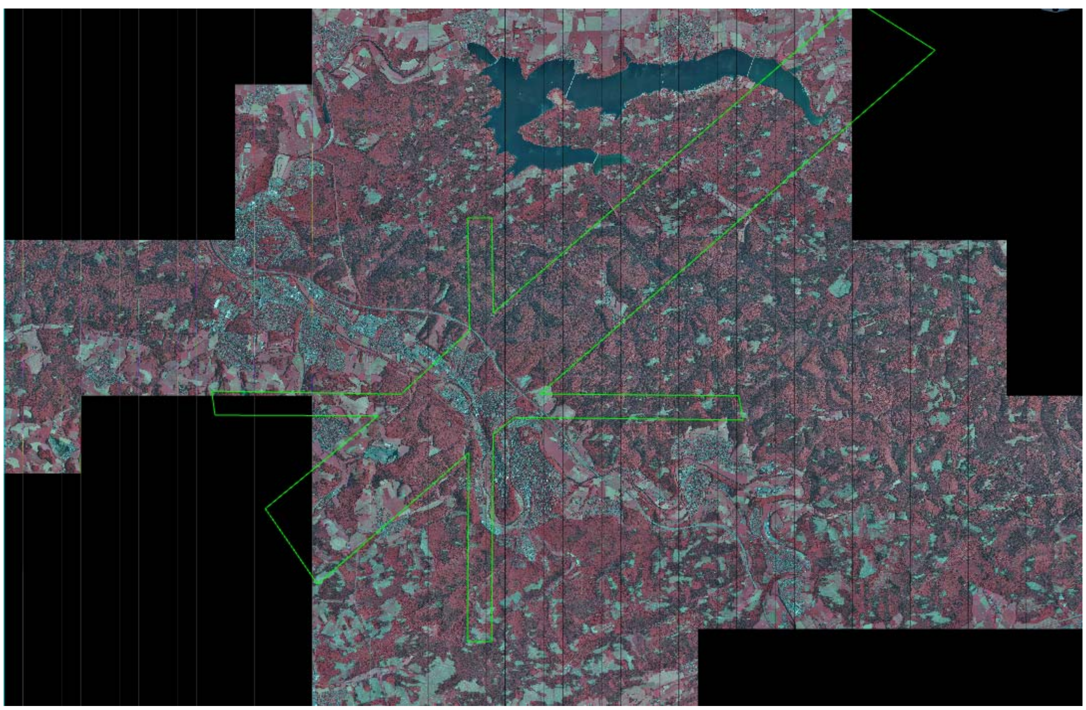

Figure 1. Study site: Overview of the CIR images. The green polygon shows the area where the hyperspectral data set is available.

For the RGB and CIR images an Ultracam X camera with a resolution of 10cm per pixel was used. The infrared band was acquired at a wavelength of around 840nm. In addition satellite data from the SPOT satellite provided red, green, near-infrared band and also a short wavelength infrared band. This data set has a resolution of 10m per pixel for the red, green and near infrared bands and 20m per pixel for the short wavelength infrared band. Based on a similar data set a decision tree based classification algorithm was developed as described in (Rossmann, 2009)

---

* The Virtual Forest is supported by the State of North Rhine Westphalia (NRW), Germany, the forest administration of North Rhine Westphalia and the European Union (Europäischer Fond für regionale Entwicklung - EFRE)

To evaluate this data set and to gain additional spectral information a hyperspectral data set was recorded with a AISA HAWK hyperspectral sensor, which is a small airborne near and short wavelength infrared sensor. The used configuration provided 320 pixels per scan line and 235 spectral bands in the range of 975nm to 2449nm with a band width of 6.3nm for each spectral band. The spatial resolution of the final product is 1.5m per pixel. This data set covers an area of 5 strips and two additional transverse strips acquired at an altitude of 1620m and with a total amount of about 80 km².

As ground truth data for the analysis 7 forest stands were measured in the field. These forest stands are shown in Figure 2. Each tree in the forest stands was measured with its position, tree species, height and additional forest parameters.

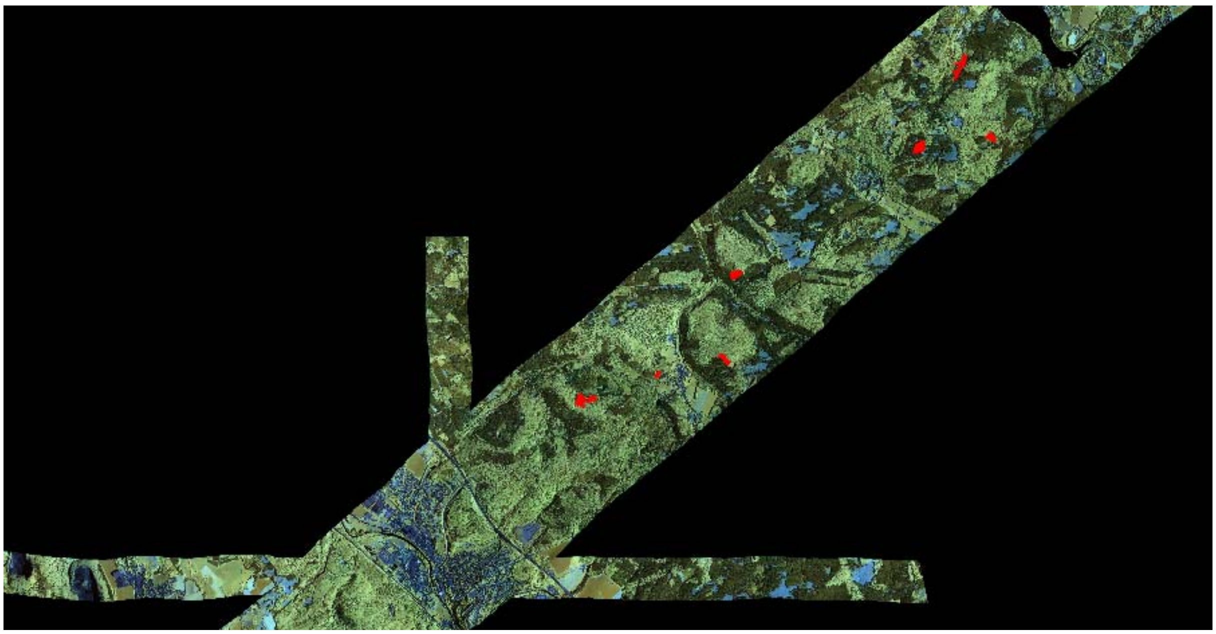

Figure 2. Hyperspectral data set. Bands 1, 50 and 100 displayed as red green and blue. The red areas are the forest stands where the reference data was measured.

The most important tree species in the federal state of North Rhine-Westphalia are oak (quercus robur), beech (fagus sylvatica), pine (pinus), european larch (larix decidua), spruce (picea), douglas fir (pseudotsuga menziesii) and poplar (populus) whereof the latter is not present in our data set. Other deciduous tree species which do occur in the designated area are less significant in the context of forestry and therefore were not part of our study.

MOTIVATION AND METHODS

In our study we omitted the lower half of the spectra for each tree species to avoid the influence of shadow pixels in our analysis. Therefore for each of the data sources, that is RGB and CIR images, SPOT satellite data and the hyperspectral data set, a near-infrared channel was chosen - as near infrared has the highest reflectance – and the samples were sorted regarding their reflectance. For each data source $1/6^{th}$ of the points was then neglected resulting in a total amount of ½ of the samples per tree species being ignored in the further analysis.

Shadow pixels occur despite georeferencing due to measurement errors in the field, understorey trees that are measured in the field but are not visible from above the canopy and due to displacement of the tree crowns in respect to the position of the roots. The remaining samples sum up to 1244 single trees of 6 tree species whereof 305 belong to Douglas fir, 120 to European larch, 249 to spruce, 6 to pine, 542 to beech and 22 to oak. Although the number of samples for pine is small and therefore might not be representative, we kept it in our study to get an estimate of the position of the spectrum of pine relative to the other spectra. The mean and the standard deviation were calculated and are displayed in the Figures below where the standard deviation is represented by error bars.

This study is based on the desire to develop a decision tree based classification for tree species discrimination. We found that tree species can be distinguished based on few specific features. However, these features differ based on the species that are to be discriminated. Depending on the used algorithms, taking insignificant features into account can worsen the classification result.

Decision trees do not base their ruling on all the features in each step but instead only the relevant features are taken into account. Therefore decision trees are a suitable approach for tree species classification. During the first tests we found that RGB and color infrared images are insufficient to discriminate the most important species in North Rhine-Westphalian forestry.

AIRBORNE IMAGE DATA AND SPOT SATELLITE DATA

Based on our first data set containing airborne images and color infrared images combined with SPOT satellite data we extracted the spectral characteristics as shown in Figure 3. The error bars in this figure represent the standard deviation of the spectral values for each tree species. Although the spectra of pine, beech and spruce are quite distinct they cannot be classified as the within-species variability is high.

The spectrum that can be identified best is pine in the red band but, even there at most a little more than 50% can be classified correctly, depending on the threshold that is used. Additionally a few faulty classifications of Douglas fir, European larch, oak, beech and even spruce will occur. Classification in the green and blue band would be worse.

The spectral values of beech differ from the other tree species except pine in the infrared band but the standard deviation especially of Douglas fir and oak but also those of the other species overlap. Furthermore a discrimination between beech and pine would still be needed.

The mean spectrum of spruce is below the spectra of the other tree species in most bands. However, the overlap with the standard deviation of the other species is much worse than for the aforementioned pine and beech.

Based on this spectral properties we cannot confirm the findings of (Key, 2001) that the blue band is the best individual band for tree species discrimination in our data. This might be due to the fact that our test area included other tree species than the area used in their study. As we will see below, tree species spectra differ from each other at different wavelengths and combinations of wavelengths.

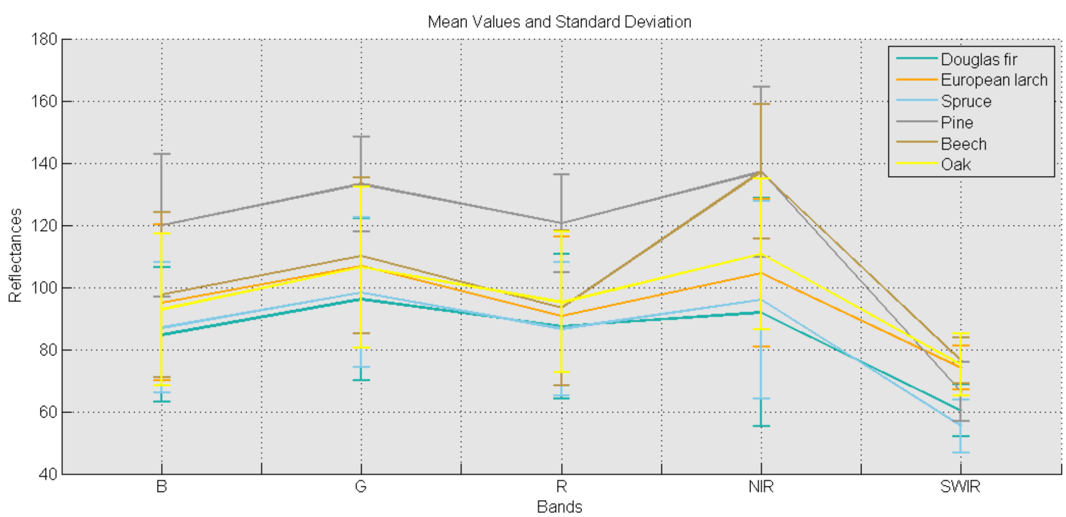

Figure 3. Mean values of the blue, green, red, near infrared and short wavelength infrared bands.

A prominent feature in Figure 3 is the slope that arises for the beech spectrum between the red and the infrared band when the spectral values are connected by lines. Difference bands can decrease the variability which can occur due to lighting changes or even changes in the water content of the leaves (Stoffels, 2009). They also manage to bring forward certain features. Therefore difference bands and in addition to that ratio bands were calculated.

The spectral characteristics and their variations have been displayed by the use of mean and standard deviation so far. To ensure that this was a valid representation the Kolmogorov-Smirnov-Test as described in (Press, 1988) was performed on several bands comparing the distribution of the samples to the Gaussian distribution. The estimated cumulative distribution function $S_N(x)$ of the data points is compared to the cumulative distribution function of the Gaussian distribution P(x) and the maximum value of the absolute difference between the two functions is calculated as in (1).

$$D = \max_{-\infty < x < \infty} \left| S_N(x) - P(x) \right| \tag{1}$$

The probability of similarity is given by equation (2)

$$p = Q_{KS}\left(\left[\sqrt{N} + 0.12 + \frac{0.11}{\sqrt{N}} \right] D \right) \tag{2}$$

with N denoting the number of samples and Q_{KS} the significance as given in equation (3).

$$Q_{KS}(\lambda) = 2\sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2\lambda^2} \tag{3}$$

Difference and ratio bands are visually still similar to the Gaussian distribution but did not pass the test. Therefore, instead of the mean and standard deviation the 0.1587 quantile, the 0.5 quantile, also known as the median, and the 0.8413 quantile are used for the visual representation of difference and ratio bands. The connecting lines between the individual difference and ratio bands for each tree species do not have any meaning. They do not approximate the spectrum as it was the case in Figure 3. They are used for a better visible representation only.

Three distinct characteristics can be observed in the first chart of Figure 4 which shows the calculated difference bands. Beech is separated from the other tree species in the differences of the near infrared band (NIR) with the red, green and blue bands. The difference in the NIR-G band is slightly larger than in the other 2 bands with only small overlaps with the standard deviations of pine and oak. European larch is set apart from the rest of the species in the difference between the short wavelength infrared band (SWIR) and the near infrared band and Douglas fir and spruce are separated in the SWIR-R and SWIR-G bands with only a small overlap with the mean error of oak. The representation of pine in the difference and ratio bands shows an extremely large variation which is probably due to the small number of samples and therefore has to be revised.

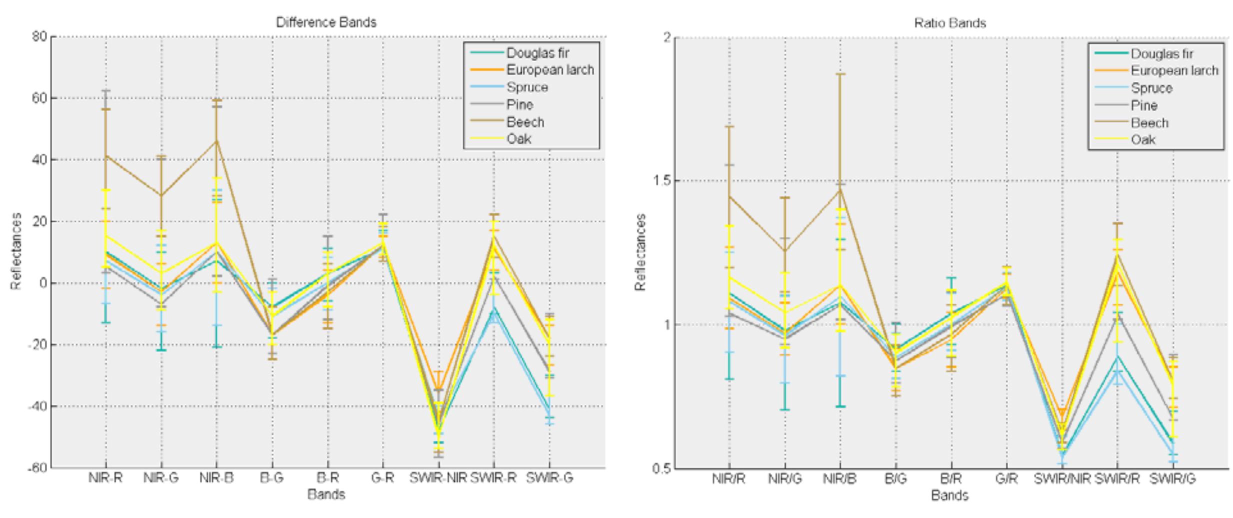

Figure 4. Difference and ratio bands.

In addition to the difference bands, ratio bands were calculated as an additional simple mean to show the value of band combinations while reducing the effect of lighting or water content of the leaves and other effects resulting in a change in the overall reflectance of the spectra. The ratio bands are shown in the second chart of Figure 4.

In the ratio bands similar characteristics can be observed. However the attributes slightly differ from the ones in the difference bands, as for example spruce and Douglas fir differ more in the SWIR/R and SWIR/G bands than they did in the SWIR-R and the SWIR-G bands.

The bands used so far are red, green, blue, one near infrared band and one short wavelength infrared band. All of the available bands show relevant information content, but in some cases the overlaps are substantial and more information is needed for a reliable discrimination. Although very high resolution hyperspectral data is not affordable for large scale forest inventories, the purpose is to find relevant bands which can be added to an ordinary sensor setup for airborne data acquisition e.g. as single band sensors.

Furthermore the information on interesting bands for tree species classification can be used to find a suitable sensor or data provider.

HYPERSPECTRAL DATA

As tree species classification is a difficult problem, with highly overlapping characteristics, we were searching for additional features that would help with the discrimination. We did find some literature on hyperspectral data which gave information on the general shape of tree species spectra in the near and short wavelength infrared areas (Pinard, 2003) and which proposed wavelengths that had proven useful in other studies (Key, 2001 and van Aardt, 2000). However they neither gave information on the overlap of the spectra for certain tree species, nor an overview over all the bands which might be useful.

Very few groups are in the position to have a full hyperspectral data set on a large test area. Most have to choose from the available multispectral data sources. Therefore it is important to know which bands can give additional information.

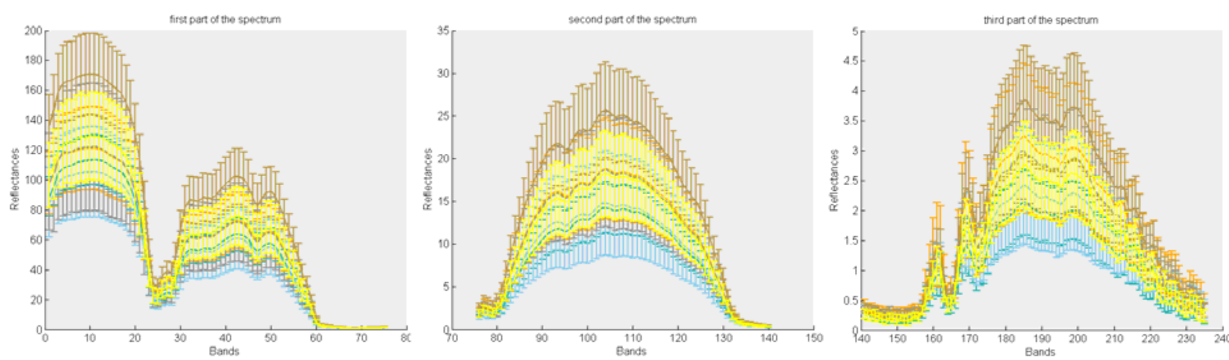

Figure 5. The AISA spectrum for the 6 tree species including one standard deviation displayed as error bars.

The spectra of the 6 tree species used in this study have a high overlap in our hyperspectral data set as shown in Figure 5. Methods like penalized discriminant analysis (PDA) as described in (Hastie, 1994) and used in (Yu, 1998) produce linear combinations that show how the components of the predictor vector contribute to the discrimination rule. The goal in this work is not to find discrimination rules, but to give an overview of the possible contributions of hyperspectral bands.

Therefore one of the simplest linear combinations – the difference – is used and in addition ratio bands were calculated again. Hyperspectral data is noisy and has small peaks and valleys in the spectrum. Therefore a 3 point mean was calculated to decrease the noise present in the collected samples without blurring the spectral information. From the whole spectrum 22 bands at those wavelengths where local maxima or minima occurred were selected. Local minima and maxima give a good representation of the overall spectrum, as the spectra do not change that much from band to band. For example taking the first local maximum at band 11 and taking a look at the area from band 4 to band 16 in Figure 5 one can see that the reflectance changes only a little and about the same for all tree species. Therefore the differences between the bands in this area amount to almost zero and the characteristic features are not distinct due to the large variation of the reflectance. The chosen bands and wavelengths are given in Table 1.

Table 1. Bands and wavelengths containing local extrema

Band Number	Wavelength	Band Number	Wavelength
11	1039.2 nm	107	1642.7 nm
26	1132.7 nm	109	1655.3 nm
43	1239.7 nm	147	1894.5 nm
48	1271.2 nm	152	1925.0 nm
51	1290.1 nm	155	1944.9 nm
70	1409.7 nm	186	2140.0 nm
92	1541.9 nm	192	2177.8 nm
96	1573.4 nm	195	2196.7 nm
100	1598.6 nm	196	2203.0 nm
101	1604.9 nm	200	2228.2 nm
105	1630.0 nm	235	2448.5 nm

The differences of the first local maximum, band 11, with the other local extremes are displayed in Figure 6. Those are the first 21 difference bands.

The other difference bands have similar characteristics to those shown in Figure 6 except for a scaling factor. The variations relative to the absolute values are almost the same as in the bands shown above. For the error bars of the difference bands there is a small overlap between the characteristics of beech and the other species. For all the other species almost the whole range of the variation overlaps with other tree species. Although European larch, pine and spruce are set off from oak, beech and Douglas fir in the peaking ratio bands in the second chard of Figure 6, there still is an overlap. Furthermore the overlap between European larch, pine and spruce as well as the overlap between beech, oak and Douglas fir is crucial.

It can be seen that the relative differences do not change much in the difference bands. The value changes, but the overlap of the variation for the tree species for the difference bands stays almost the same relative to the overall value of the reflectance and the variations. The variation – that is the 0.1587 quantile and the 0.8413 quantile – changes in a similar manner, as does the absolute value of the difference bands. Again beech is set off from the other spectra more clearly in the difference bands than in the spectral reflectances. Overall the difference bands yield almost the same information over the whole near infrared and short wavelength infrared region – despite a scaling factor. The analysis indicates that in general the data in the visible part of the spectrum provides more useful information for tree species classification.

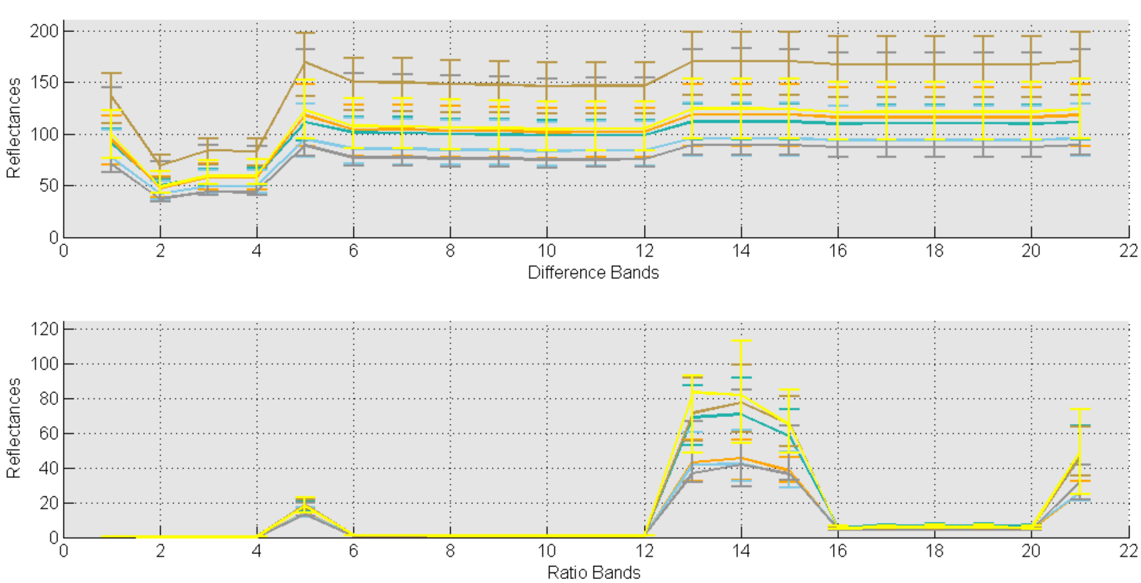

Figure 6. First 21 AISA difference and ratio bands

The second chart of Figure 6 shows the ratio bands calculated as a combination of band 11 with all the other bands. The information content is very little in most of the bands and the within-species variation outweighs the between-species variation. For the band combinations 13 to 15, which are calculated with the 147th, 152nd and 155th band, the spectra of European larch, spruce and pine are set apart from Douglas fir, beech and oak. The same characteristics can be found in some of the other ratio bands containing those 3 particular bands.

However based on the visible data spruce and Douglas fir as well as European larch, pine and oak seem to be difficult to separate. Therefore we separated those tree species spectra from the rest and analyzed them individually. The difference bands did not yield additional information for the discrimination of Douglas fir and spruce. The ratio band showed several significant divergences between the spectra for the ratios calculated with the 147th, 152nd and 155th bands. The most significant ratios were calculated with the 26th band and are shown in Figure 7. The spectra of European larch, oak and pine show significant differences for the same ratio bands, as did the Douglas fir and spruce spectra and are shown in the second chart of Figure 7.

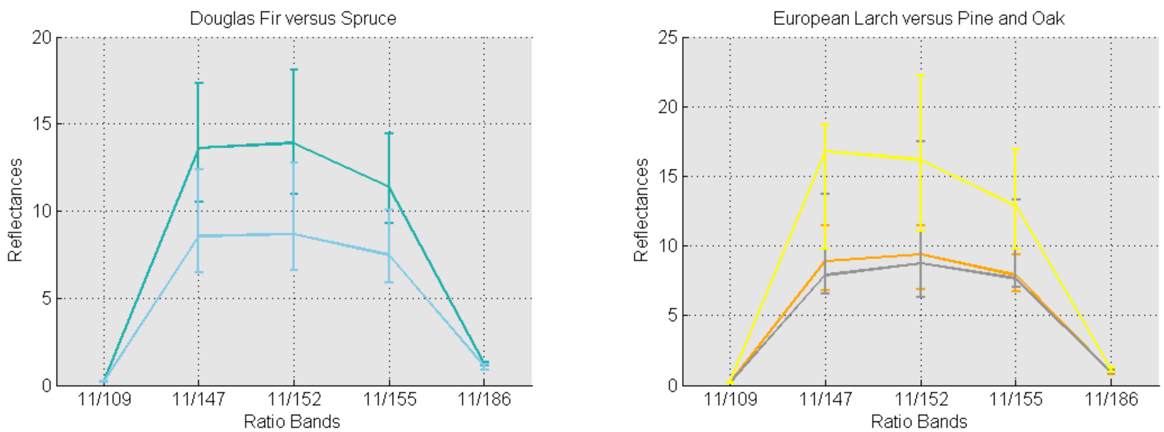

Figure 7. Douglas fir, Spruce, European larch, pine and oak in the AISA ratio bands.

CONCLUSION

Hyperspectral data sets have high information content but tend to be noisy. Therefore a 3-point mean was calculated for the spectral bands. Furthermore the variances in the near infrared and short wavelength infrared are large and highly overlap for the 6 tree species that are used in this study, oak, beech, Douglas fir, spruce, European larch and pine. The samples of pine might not be representative due to the small number of samples available for this tree species.

We showed that difference and ratio bands can be useful for tree species discrimination, as they can reduce the effect of changes in the overall reflectance of a spectral curve. A number of difference and ratio bands were calculated to give an overview of the information that can be gained from the near infrared and short wavelength infrared regions. Although the data from the visible and the first part of the near infrared region showed to be more useful for tree species discrimination the hyperspectral data also showed that short wavelength infrared regions contains helpful information. The area around 1550nm helps in the separation of Douglas fir and spruce from the other tree species, and the region at 1900nm in combination with the near infrared regions at about 1100nm help in the distinction between Douglas fir versus spruce and larch versus oak. An alternative source for a band in the short wavelength infrared region might be airborne laser scanner data, which is used in most forest inventories based on remote sensing data, to estimate tree heights and other forest parameters. Several common lasers used for airborne data acquisition operate with a wavelength in the short wavelength infrared area. Evaluation of laser scanner data not only for the distance of the reflection but also for the intensity of the reflection might provide a useful additional band for classification in forest areas.

From the available bands combinations of the red, green and near infrared bands with the short wavelength infrared bands yielded crucial information for the discrimination of some of the tree species.

OUTLOOK

The information given in this study can help with the search for a suitable data source for tree species classification. Furthermore the information on difference and ratio bands can also be used to improve existing classification algorithms and to develop new methods. No data on the red edge region was available for this study. Therefore, we cannot give a statement as to whether it is a useful band for tree species classification as stated in (Gong, 1997). The next steps will be to use the information given in this paper for the development of a new classification algorithm and to test the results against widely used methods like the kNN Algorithm as described in (Tomppo, 2007).

REFERENCES

Gong, P., R. Pu and B. Yu, 1997. Conifer Species Recognition: An Exploratory Analysis of In Situ Hyperspectral Data, *Remote Sensing of Environment*, *62*:189-200.

Hastie, T., A. Buja, & R. Tibshirani,1995. Penalized discriminant analysis. Annals of Statistics 23, 73–102.

Key, T., T.A. Warner, J.B. McGraw and M.A. Fajvan, 2001. A Comparison of Multispectral and Multitemporal Information in High Spatial Resolution Imagery for Classification of Individual Tree Species in a Temperate Hardwood Forest. *Remote Sensing of Environment 75*:100-112.

Kirkman, T.W. (1996) Statistists to Use. http://www.physics.csbsju.edu/stats/ (17.02.2010)

Pinard, V. and A. Bannari, 2003. Spectroradiometric analysis in a hyperspectral use perspective to discriminate between forest species, *Proceedings of the IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, 7(7):4301-4303.

Rossmann, J. and P.M. Krahwinkler, 2009. Tree Species Classification and Forest Stand Delineation Based on Remote Sensing Data – Large Scale Monitoring of Biodiversity in the Forest, *Proceedings of the International Symposium on Remote Sensing of Environment (ISRSE) 33*, Stresa, Italy.

Stoffels, J. 2009.: Einsatz einer lokal adaptiven Klassifikationsstrategie zur satellitengestützten Waldinventur in einem heterogenen Mittelgebirgsraum. *Thesis*, Trier, Germany.

Tomppo, E., H. Olsson, G. Ståhl, M. Nielsson, O. Hagner and M. Katila, 2008. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment* 112, 1982-1999.

Van Aardt, J.A, 2000: Spectral Separability among Six Southern Tree Species. *Thesis*, Blacksburg, USA

Yu, B., M. Ostland, P. Gong and R. Pu, 1999. Penalized discriminant analysis of in situ hyperspectral data for conifer species recognition. *IEEE Transactions on Geoscience and Remote Sensing*, 37, 2569-2577.