ASSESSMENT OF HYPERSPECTRAL DATA ANALYSIS METHODS TO CLASSIFY TREE SPECIES

Gints Priedītis, Ingus Šmits, Salvis Daģis, Līga Paura, Jānis Krūmiņš, Dagnis Dubrovskis
Latvia University of Agriculture
gints.prieditis@llu.lv

Abstract
One of the most challenging issues in forest inventory based on remote sensing data is identification of tree species. Hyperspectral remote sensing data provides information which considerably facilitates tree species recognition. The objective of the research is to evaluate different hyperspectral data analysis methods to classify tree species in Latvian forest conditions. The study site is a forest in the central part of Latvia, Jelgava district (56º39' N, 23º47' E). The area consists of a mixed coniferous and deciduous forest. During research 598 trees were analyzed in 70 sample plots. Remote sensing data are 64 hyperspectral bands in the 400 - 970 nm spectral range. Two different classification techniques: linear discriminant analysis (LDA) and artificial neural networks (ANNs) were used. In LDA species classification was done by stepwise and using principal components of hyperspectral bands. In stepwise LDA 18 hyperspectral bands were used. LDA using principal components and ANNs used all 64 hyperspectral bands. The best results show stepwise LDA where 82.4% of the data were correctly classified. Scots pine was classified 94.8%, Norway spruce 83.5%, Silver birch 77%, European aspen 71.4% and Black alder 56.3%. Classification with ANN's best results showed for Scots pine, Norway spruce and Silver birch – respectively 81%, 84%, 86%. With LDA using principal components Scots pine's classification showed best results with 85.1% correctly classified trees.
Key words: Forest inventory, hyperspectral remote sensing, tree species identification, linear discriminant analysis.

Introduction

The objective of forest inventory is to provide forest owners with the necessary information dealing with forest management planning, documentation of economic activities and for stating the value. The quality of the information is very topical in forestry industry because it leads to a precision of evaluating the condition of forest resources and helps to take the most effective forest management planning decisions.

Forest inventory methods that are based on remote sensing data are economic and less time consuming (Hyyppä et al., 2008). Up until now, the studies in Latvia and in the world show that separate stand characteristics can be identified with a high precision. At the same time it has been pointed out that results are substantially influenced by the structures of remote sensing data, additional information about the stand, and data processing methods being used (Korpela et al., 2006).

Main problems impeding the development of this technology in the conditions of Latvia are related to the fact that data processing methods cannot fully evaluate and automatically identify all economically important tree species (Priedītis et al., 2013).

Promising results in identification of tree species are obtained with the help of hyperspectral remote sensing that allows to analyse hundreds of spectral bands. Hyperspectral remote sensing data provides information of many spectral bands. Most frequently, to obtain optimal results, only a part of proposed spectral bands is used (Ghiyamat and Shafri, 2008; Thenkabail et al., 2004). Several scientists consider that approximately 90% of hypesrspectral remote sensing data are unnecessary and can even interfere with the identification of tree species (Thenkabail et al., 2004). Latest scientific improvements in hyperspectral image acquisition have allowed to improve the spectral and spatial resolution, which allows to obtain even larger amount of information, and as a result, it allows to separate and analyze very narrow spectral bands (Thenkabail et al., 2014).

Various statistical approaches are used in most of data processing methods when analysing separate spectral bands (Masaitis and Mozgeris, 2013; Thenkabail et al., 2004), their combinations (Agapiou et al., 2012; Agapiou et al., 2012; Blackburn, 2007; Thenkabail et al., 2004; Thenkabail et al., 2014), or indices (Adam et al., 2010; Haboudane et al., 2004; Thenkabail et al., 2004; Thenkabail et al., 2014). Different formulas are used to calculate indices, using information only from approximately 30 spectral bands (Blackburn, 2002; Thenkabail et al., 2002), the used spectral bands in various studies are observed as different. Classification methods that reduce the high range of data and process only that part which is characterized by tree species, are also frequently used (Dinuls et al., 2011; Dinuls et al., 2012; Waser et al., 2010). Reflection intensity, which is analysed in various spectral bands, can be dependent on many internal and external factors (Adam et al., 2010) on a species level, thus influencing identification and classification results of the tree species. Different studies have shown that, despite the large number of spectra, spectral measurements tend to be very similar for different tree species (Waser et al., 2010; Wen et al., 2008).

The use of hyperspectral remote sensing data in forest inventory for classification of tree species is

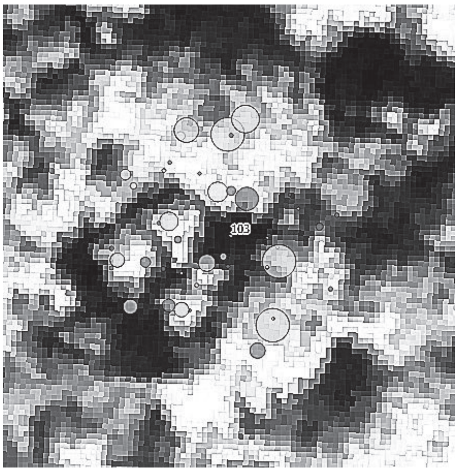

Figure 1. Aerial photography with actual on-site situation of the sample plot.

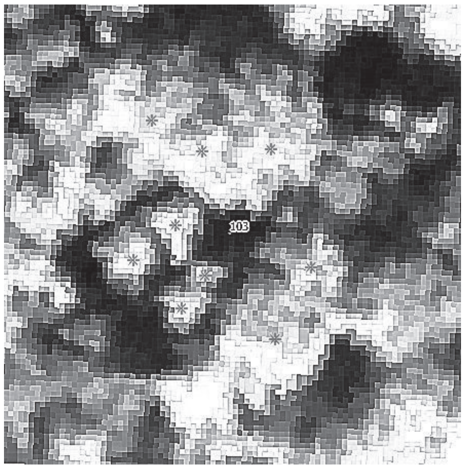

Figure 2. Aerial photography with adjusted situation of the sample plot and marked sample trees.

extensively researched in various types of forests. The structural characteristics of a stand is one of the reasons which determine the precision level of results, applicable data processing methods, and also most suitable spectral bands. When hyperspectral remote sensing data is analysed, it is possible to obtain information about biochemical (Curran et al., 2001) and biophysical (Darvishdaeh, 2008; Zhang et al., 2013) characteristics of a plant, which in several cases can help to identify the tree species.

The objective of the research is to evaluate different hyperspectral data analysis methods to classify tree species in Latvian forest conditions.

Materials and Methods

The study site was a forest in the central part of Latvia, Jelgava district (56º39’ N, 23º47’ E). Totally 70 sample plots (0.05 ha) were established. The area consists of a mixed coniferous and deciduous forest of different age, high density, complex structure, various components, composition and soil conditions. Represented species are Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* (L.) H.Karst), Silver birch (*Betula pendula* Roth), Black alder (*Alnus glutinos* L.), and European aspen (*Populus trémula* L.).

All trees with a diameter at breast height DBH of more than 5 cm were measured and for each tree coordinates, its species, height, DBH, crown width and length were recorded. Differentially corrected Global Positioning System measurements were used to determine the position of each plot centre. The accuracy of the positioning was less than 1 meter.

Individual trees from the sample plots were used in the study. Data processing consisted of manually selecting trees with recognizable tree crowns in the images. Tree centres were adjusted by putting them in the accurate position according to the situation in aerial photography (Figure 1 and Figure 2).

It was performed with an aim to exclude miscalculations that could appear when inaccurate tree centres would be determined by using automated approach. Study includes sample plot measurements for 598 trees - Scots pine (209 trees), Norway spruce (133 trees), Silver birch (147 trees), Black alder (47 trees), and European aspen (62 trees).

Remote sensing data were obtained using a specialized aircraft (Pilatus PC-6), which is equipped with a high-performance airborne VNIR pushbroom hyperspectral system (AisaEAGLE). System acquires full, high quality hyperspectral data up to 488 spectral channels with 1024 swath pixels and high image rates. The study area was flown at 1000 m altitude. Data was recorded in 64 hyperspectral bands in the 400 - 970 nm spectral range, spectral resolution was 3.3 nm, ground resolution 0.5 m, a field of view from 17.36° to - 18.68°. Flight was done on 27 October, 2014.

Two different classification techniques were used in the paper: linear discriminant analysis (LDA) and artificial neural networks (ANNs). Data of 64 hyperspectral bands was analysed to classify and predict five trees species: Scots pine, Norway spruce, Silver birch, Black alder and European aspen. Before the tests, the outliers were excluded from the data set and each predictor variable was normally distributed. Most of hyperspectral bands are correlated with each other, therefore for the LDA application from the large number of bands were selected uncorrelated bands. The number of hyperspectral bands was reduced by stepwise LDA and factor analysis (PCA). According to stepwise LDA, 18 hyperspectral bands contributed to the difference between the groups and these bands were included in the discriminant function. According to PCA, 64 hyperspectral bands were combined to

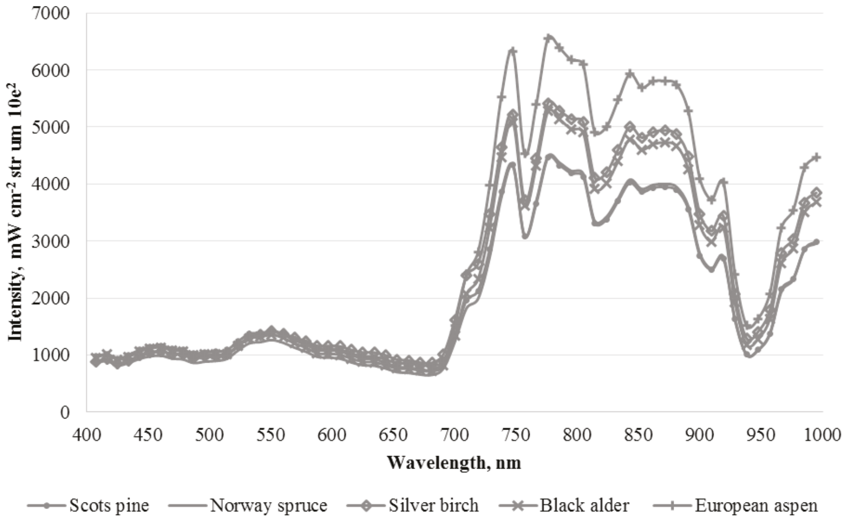

Figure 3. Average reflection intensity in different spectral bands according to the tree species.

four factors (PC) and new factors (PC) were used in LDA to classify tree species.

ANNs training was performed with 100 Scots pine, silver birch and Norway spruce trees, but for European aspen ANNs training was done with 42 trees and for black alder with 27 trees. For the test data, 20 trees of each species, which would not overlap with the training set of data, were selected. For each test, the set of data was randomised and only the specified tree count was selected. Tree identification test was repeated 4 times. For the ANNs, 64 spectra were used in the input, and 5 species in the output.

Results and Discussion

Differences between trees species means were analysed by ANOVA and differences were observed for all predictor variables or 64 hyperspectral bands. Mean differences between interested tree species in red and near infrared hyperspectral bands are larger and these bands may be good discriminators for separation of groups (Figure 3).

Literature analysis shows that spectral features of tree species during the year vary by seasons (Masaitis and Mozgeris, 2013). Our research results about spectral ranges with largest differences between tree species are similar to the findings in the studies of other scientists.

Species classification by stepwise LDA

According to stepwise LDA, the 18 hyperspectral bands were significant and were used to classify five tree species. The intracorrelations between blue and red, and between violet and red infrared hyperspectral bands are low. The 18 bands discriminate functions revealed significant associations between groups and predictors. Four canonical discriminant functions were used in the analysis (Table 1) and the three models explain 95.4% of the variation in the grouping variable – tree species. The first discriminant function explains 49.8% of the possible differences among the five tree species groups; whereas bands more than 700 nm were the stronger predictors and 400-600nm bands were with less predictor with smaller standardized canonical discriminant function coefficients.

The second discriminant function explains 29.4% of the possible differences among the species. 8 hyperspectral bands in range 400-600nm have a higher loading on the second discriminant function and can be labelled “visible light spectres”. The first discriminant function is statistically significant and with this function it is possible to classify two groups’ – differences among the Scots pine, Norway spruce and Silver birch, Black alder, European aspen (Figure 4). The second discriminant function can be used to classify Scots pine from Norway spruce and the 3rd Silver birch from European aspen.

Table 1

Eigenvalues, variance and canonical correlation of 18 bands discriminant functions

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.2	49.8	49.8	0.830
2	1.3	29.4	79.2	0.753
3	0.7	16.2	95.4	0.647
4	0.2	4.6	100.0	0.412

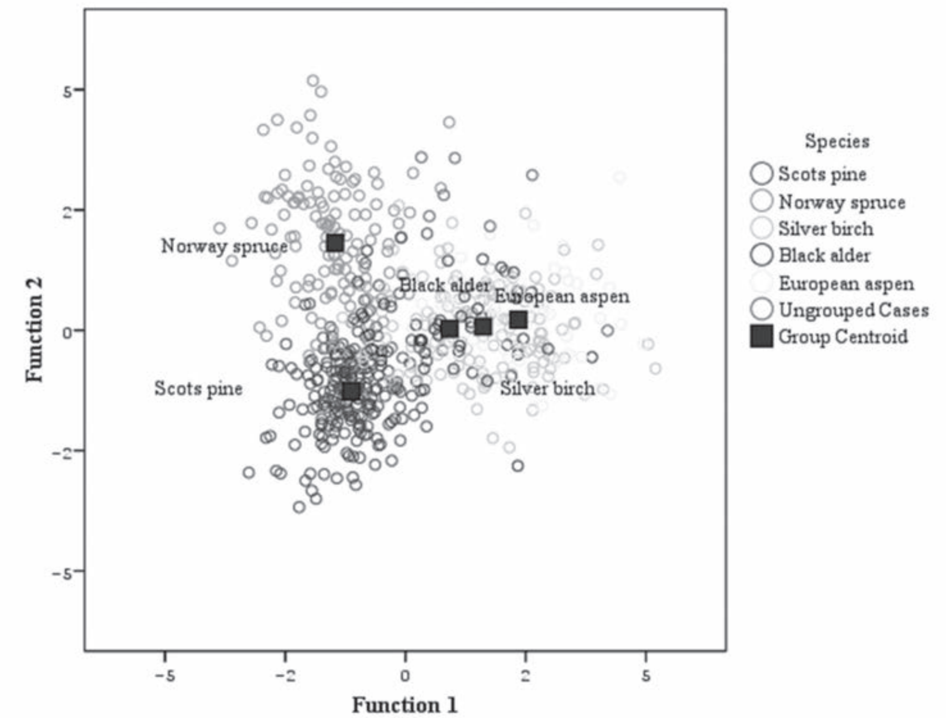

Figure 4. Plot of five tree species groups in discriminant space.

The classification results of tree species are shown in Table 2 and Figure 3. 82.4% of the data (trees) were correctly classified. Scots pine and Norway spruce were classified with a slightly better accuracy 94.8% and 83.5%, respectively, than other tree species. 14.3% of Norway spruce and 14.6% of Black alder were classified as Scots pine. 15.9% of European aspen were classified as Silver birch. The cross validated classification showed that 81.1% of trees were correctly classified.

Species classification by stepwise LDA is used for identification of tree species in varied structure forest stands. Literature analysis shows that even if 15 tree species are classified, this method can produce an overall classification accuracy of 86% for dense crowns (Alonzo et al., 2013).

Species classification by LDA using principal components of hyperspectral bands

The hardest part of hyperspectral data evaluation process is to decide which bands are the most valuable for given task. One of the common solutions for such a problem is to use the principal component analysis, which investigates if it is possible to represent variation between data samples using a smaller number of variables (usually called principal components). Principal components show the amount of variance described by each of principal components which is

Table 2

Tree species classification results of the total number of cases obtained with 18 hyperspectral bands, %

	Tree species	Scots pine	Norway spruce	Silver birch	Black alder	European aspen
Original	Scots pine	**94.8**	3.3	1.9	-	-
	Norway spruce	14.3	**83.5**	2.3	-	-
	Silver birch	10.1	2.7	**77.0**	6.1	4.1
	Black alder	14.6	6.3	8.3	**56.3**	14.6
	European aspen	3.2	1.6	15.9	7.9	**71.4**
Cross validation	Scots pine	**94.3**	3.3	1.9	.5	-
	Norway spruce	14.3	**83.5**	2.3	-	-
	Silver birch	10.1	4.1	**75.0**	6.8	4.1
	Black alder	14.6	8.3	10.4	**52.1**	14.6
	European aspen	3.2	1.6	15.9	11.1	**68.3**

Table 3

Tree species classification results of the total number of cases obtained with PC of hyperspectral bands, %

Tree species	European aspen	Silver birch	Norway spruce	Black alder	Scots pine
European aspen	**54.1**	24.3	8.1	5.4	8.1
Silver birch	6.7	**72.7**	4.6	-	16.4
Norway spruce	1.1	5.3	**50.5**	5.3	37.9
Black alder	10.5	21.1	-	**15.8**	52.6
Scots pine	0.6	4.6	9.7	-	**85.1**

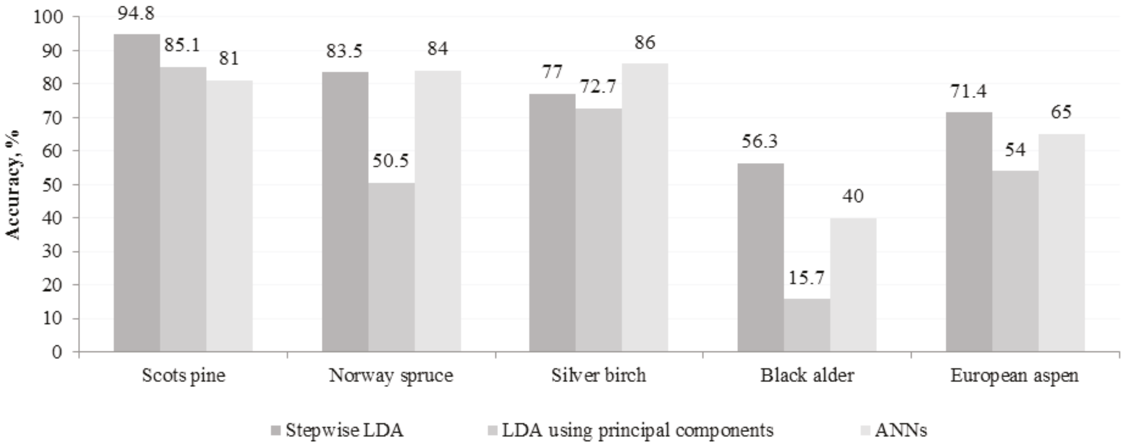

Figure 5. Summary of tree species classification methods.

acquired by transforming all 64 bands of hyperspectral data. The biggest amount of variances is described by the first component and four components describe in total 0.9961% of data variance. LDA model which uses transformed data of spectral bands to classify tree species were created. Tree species classification results obtained with PC are shown in Table 3 and Figure 5.

Scots pine and Silver birch were classified with accuracy of 85.1% and 72.7%. Black alder was classified with accuracy of 15.8%.

The PC has been used as a method for best band selection in many studies and is a widely used method (Agapiou et al., 2012; Ghiyamat and Shafri, 2008; Masaitis and Mozgeris, 2013). The use of this method in our research shows poor results, which could be due to the use of all available bands in the data processing. In scientific literature, some authors suggest using a smaller set of selected bands which could generate more accurate results than the whole set of spectral bands (Adam et al., 2010; Krahwinkler and Rossmann, 2010).

Species classification by ANNs

Results of tree species classification with ANNs in the studies of different authors are controversial, because precision of the results depends on many parameters and in most cases these parameters are unique within the certain trial. Precision of this method is basically influenced by the amount of input and training data.

In the research best results of tree species classification with ANNs were achieved for Scots pine, Norway spruce and Silver birch – respectively 81%, 84%, 86% of trees were classified correctly. But for Black alder and European aspen only 40% and 65% of trees were classified correctly (Figure 5), which could be explained due to the small number of training data – 27 for Black alder and 42 for European aspen.

Summary of Tree species classification methods

In the research of different authors the tree species identification using hyperspectral remote sensing data is performed with the precision from 50 till even 97%, which depends on tree species and used data processing methods (Adam et al., 2010; Chan and Paelinckx, 2008). Several studies show a 90% precision for separate tree species (Agapiou et al., 2012; Boschetti et al., 2007; Clark et al., 2012). In the ideal conditions, tree species can be identified with precision, which is higher than 95 % (Dinuls et al., 2011; Dinuls et al., 2012). Hyperspectral remote sensing data in combination with Lidar data is used to increase and obtain more precise results (Blackburn, 2002; Dinuls et al., 2012).

Conclusions

1. The presented classification methods LDA and ANNs in Latvian forest conditions are providing classification accuracy in range of 40-95%.
2. LDA using 18 hyperspectral bands show the highest tree species classification, where on average 82.4% of the trees were correctly classified. The Scot pine and Norway spruce species had higher classification results.
3. ANNs classifier provided similar classification results as LDA. The Scot pine, Norway spruce and Silver birch species had higher classification results. Only 40% of Norway spruce was correctly classified.
4. LDA classification using principal component produced very weak results and the number of classification trees had decreased.
5. The wavebands which showed the largest spectral differences about interested species in autumn are red and near infrared spectral zones.
6. Combinations with other remote sensing data sources could improve tree species classification results, as well as the combinations of data processing methods.
7. Latvian forest conditions are difficult for all remote sensing methods used mainly for mixed deciduous and coniferous spaces with high density and homogeneous crown.

Acknowledgement

This research was funded by the **European Regional Development Fund project** “Technology for remote forest inventory using hyper spectral and LiDAR data” (State Education Development Agency, Latvia University of Agriculture, contract number: 2014/0050/2DP/2.1.1.1.0/14/APIA/VIAA/048).

References

1. Adam E., Mutanga O., Rugege D. (2010) Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation. *Wetlands Ecology and Management*, 18(3), pp. 281-296.
2. Agapiou A., Hadjimitsis D.G., Alexakis D.D. (2012) Evaluation of Broadband and Narrowband Vegetation Indices for the Identification of Archaeological Crop Marks. *Remote sensing,* 4(12), pp. 3892-3919.
3. Alonzo M., Roth K., Dar R. (2013) Identifying Santa Barbara’s urban tree species from AVIRIS imagery using canonical discriminant analysis. *Remote Sensing Letters*, Vol. 4, No. 5, pp. 513-521.
4. Blackburn G.A. (2002) Remote sensing of forest pigments using airborne imaging spectometer and LIDAR imagery. *Remote Sensing of Environment*, 82, pp. 311-321.
5. Blackburn G.A. (2007) Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany*, 58(4), pp. 855-867.
6. Boschetti M., Boschetti L., Oliver S., Casati L., Canova I. (2007) Tree species mapping with Airborne hyper-spectral MIVIS data. *International Journal of Remote Sensing*, 28(6), pp. 1251-1261.
7. Chan J.C.W., Paelinckx D. (2008) Evaluation of Random Forest and Adaboost tree-based ensemble classification andspectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112, pp. 2999-3011.
8. Clark M.L., Roberts D.A. (2012) Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. *Remote Sensing*, 4(6), pp. 1820-1855.
9. Curran P.J., Dungan J.L., Peterson D.L. (2001) Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: Testing the Kokaly and Clark methodologies, *Remote Sensing of Enivironment*, 76, pp. 349-359.
10. Darvishzadeh R. (2008). Hyperspectral remote sensing of vegitation parameters using statistical and physical models. *ITC Dissertation*. International Institute for Geo-information Science & Earth Observation, Enschede, the Netherlands, 169 p.
11. Dinuls R., Erins G., Lorencs A., Mednieks I., Sinica-Sinavskis J. (2012) Tree species identification in mixed Baltic forest using lidar and multispectral data. *Selected Topics in Applied Earth Observations and Remote Sensing.* Vol.5, pp. 594-603.
12. Dinuls R., Lorencs A., Mednieks I. (2011) Performance Comparison of Methods for Tree Species Classification in Multispectral Images. *Electronics and Electrical Engineering*, 111(5), pp. 119-122.
13. Ghiyamat A., Shafri H.Z.M. (2008) A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *International Journal of Remote Sensing,* 31(7), pp. 1837-1856.
14. Haboudane D., Miller J.R., Pattey E., Zarco-Tejada P.J., Strachan I. B. (2004) Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90, pp. 337-352.

15. Hyyppä J., Hyyppä H., Leckie D. (2008) Review of Methods of Small-footprint Airborne Laser Scanning for Extracting Forest Inventory Data in Boreal Forests. *International Journal of Remote Sensing*, 29, pp. 339-366.
16. Korpela I. (2006) Incorporation of Allometry into Single-tree Remote Sensing with LIDAR and Multiple Areal Images. Department of Forest Resource Management. University of Helsinki, Finland. p. 6.
17. Krahwinkler P., Rossmann J. (2010) Tree Species Classification Based on the Analysis of Hyperspectral Remote Sensing Data. In: *Remote Sensing for Science, Education, and Natural and Cultural Heritage.* Proceedings of the 30th EARSeL symposium, University of Oldenburg, Oldenburg, Germany, pp. 321-328.
18. Masaitis G., Mozgeris G. (2013) The influence of the growing season on the spectral reflectance properties of forest tree species. In: Treja S. and Skujniece S. (eds) Research for Rural Development. *Annual 19th ISC proceedings Vol. 2*, Latvia University of Agriculture, Lielā ielā 2, Jelgava, Latvia, pp. 20-25.
19. Priedītis G., Šmits I., Daģis S., Dubrovskis D. (2013) Treee species identification using Lidar and optical imagery. In: Treja S. and Skujniece S. (eds) Research for Rural Development. *Annual 19th ISC proceedings Vol. 2*, Latvia University of Agriculture, Lielā ielā 2, Jelgava, Latvia, pp. 34-42.
20. Thenkabail P.S., Gumma M.K., Teluguntla P., Mohammed I.A. (2014) Hyperspectral remote sensing of vegetation and agricultural crops. Photogrammetric Engineering & Remote Sensing, 80(8), pp. 697-709.
21. Thenkabail P.S., Enclona E.A., Ashton M.S., Van Der Meer B. (2004) Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, vol. 91, pp. 354-376.
22. Thenkabail P.S., Smith R.B., De-Pauw E. (2002) Evaluation of Narrowband and Broadband Vegetation Indices for Determining Optimal Hyperspectral Wavebands for Agricultural Crop Characterization. *Photogrammetric Engineering and Remote Sensing*, 68(6), pp. 607-621.
23. Waser L.T., Baltsavias E., Ginzler C., Küchler M. (2010) Semiautomatic classification of tree species by means of multi-temporal airborne digital sensor data ADS40. In: *ISPRS Technical Commission VII Symposium*. Vol. 38 - Part 7B, pp. 633-638.
24. Wen Z., Baoxin H., Linhai J., Woods M., Courville P. (2008) Automatic Forest Species Classification using Combined LIDAR Data and Optical Imagery. In: Geoscience and Remote Sensing Symposium. Vol 5 – Part 5, pp.134-137.
25. Zhang C., Kovacs J. M., Wachowiak M.P., Flores-Verdugo F. (2013) Relationship between Hyperspectral Measurements and Mangrove Leaf Nitrogen Concentrations. Remote Sensing, 5(2), pp. 891-908.