

Mapping Tree Species Using LIDAR in Mixed-Coniferous Forests

Agus Suratno*, Carl Seielstad, LLOYD Queen

National Center for Landscape Fire Analysis, College of Forestry and Conservation
The University of Montana (agus.suratno@ntsg.umt.edu, carl@ntsg.umt.edu,
lpqueen@ntsg.umt.edu; *corresponding author)

Abstract

This study evaluates the capacity of low density (< 1 return/m²) ALS data for classifying and mapping Douglas-fir, western larch, ponderosa pine, and lodgepole pine in western Montana, U.S.A. A two-staged approach is used. First, Linear Discriminant Analysis (LDA)-based classification function coefficients are generated for twelve intensity and height variables and combined in a linear equation to produce a single layer representing the interaction of intensity and height as a function of species. Second, a standard supervised Maximum Likelihood Classification (MLC) is applied using the LDA-species layer and percent canopy cover (PCC) with equal prior probabilities of class samples. Validation of the thematic accuracy of the resulting maps was conducted at a 0.04 ha resolution using data from an intensive walk-through stand inventory. The MLC produced an overall accuracy of 75% with the best results for Douglas-fir followed by lodgepole pine and ponderosa pine, while western larch was difficult to identify. However, classification accuracy is low when stands without a clear dominant species are included. Almost without exception, the classification identifies the correct mix of species within each mixed polygon but field data do not currently support validation of individual pixels within these stands.

Keywords: Species, Intensity, Percent Canopy, Stand Database, Classification

1. Introduction

Tree species data are necessary in forest inventories to support various management activities including timber production, biomass estimation, wildlife habitat delineation, and site productivity prediction. Several studies have shown that airborne laser scanning (ALS) data can be used to identify tree species in different forest ecosystems (Holmgren and Persson, 2004; Órka et al. 2007; Moffiet et al. 2005; Brandtberg et al. 2003; Brandtberg, 2007; Donoghue et al. 2007; Suratno et al. in review). Holmgren and Persson (2004) and Órka et al. (2007) used height and intensity with quadratic and linear discriminant analyses (QDA/LDA) to differentiate between deciduous and conifer trees in Scandinavian forests. Brandtberg et al. (2003) and Brandtberg, (2007) analyzed ALS-derived crown characteristics (leaf on/off) using ANOVA and LDA to show that oaks, red maples and yellow poplar in eastern hardwood forest of the United States exhibited distinct canopy structures. In Australian sub-tropical forests, Moffiet et al. (2005) demonstrated that white cypress pine and poplar box trees were separable by applying exploratory data analysis and LDA on vegetation permeability, mean height, and intensity of different return types. Most recently, a study conducted by Suratno et al. (in review) showed that combinations of means and standard deviations of normalized intensity and canopy height were useful for identifying four prominent conifer species (Douglas-fir, ponderosa pine, lodgepole pine, and western larch) in mixed forests of western Montana, U.S.A. The latter authors used ANOVA and LDA to show that Douglas-fir was distinguishable from the other species using

intensity metrics alone, but discrimination between all four species required intensity and height combinations. Discriminatory power was considerably higher at plot-level (0.04ha) than for individual trees.

The consensus of previous work is that tree species identification is achievable in a range of environments using combinations of ALS-derived structure and intensity metrics with reported accuracies ranging from 67 to 95%. However, the aforementioned studies represent isolated examples at local scale with common, but constrained methodologies. In short, the classification and mapping of tree species at landscape scale using ALS-data has not yet been demonstrated in the literature. It is noteworthy that LDA with cross-validation has been utilized by many researchers to derive species classes but not applied to landscapes, and we anticipate considerable difficulty in the application of this technique to landscapes with interspersed species because many landscape features will not share characteristics with the classes used to create the discriminant functions. In addition, it is difficult to apply LDA directly for classification using discriminant functions alone because of LDA's inability to separate functions when class centroids overlap (Jieping et al., 2006). Consequently, models generated from discriminant functions may produce high error rates when attempting to allocate features to classes (Krzanowski, 1988).

An alternative method similar to LDA is maximum likelihood classification (MLC). While LDA is a statistical estimation requiring more steps prior to using it for classification with some consequence mentioned above, MLC is a classification technique widely used in the remote sensing community. MLC is based on the probability density function to calculate the likelihood that individual features belong to classes (Lillesand and Kiefer, 2000). One weakness of MLC is that every individual is assigned to a class regardless of distance from class mean. This becomes a processing time (efficiency) issue when many variables are involved (Jia and Richards, 1994). In order to overcome such problem, several studies propose reducing the number of variables using principle components analysis, supervised classification, or multi-stage classification (Jia and Richards, 1994; Guohui et al, 1999; Jia and Richards, 2003). A model based on LDA functions would produce a similar effect.

In the research described in this paper, we exploit commonalities in metrics and methods of the previously cited work to classify four conifer tree species in an 11,300 ha study area in western Montana, U.S.A. and validate the thematic accuracy of the resulting maps at a 0.04 ha resolution. A two-staged approach is used. First, Linear Discriminant Analysis (LDA)-based classification function coefficients are generated for twelve intensity and height variables and combined in a linear equation to produce a single layer representing the interaction of intensity and height as a function of species. This procedure is a logical extension of previous analyses (Holmgren and Persson, 2004; Moffiet et al. 2005; Suratno et al. in review) in which LDA has been used to discriminate species. Second, a standard supervised Maximum Likelihood Classification is applied using the LDA-species layer and percent canopy cover (PCC) with equal prior probabilities of class samples. The rationale for including PCC is that observed differences in intensity appear to be at least as much a function of canopy closure as species, and investigation of additional variables such as stem density confirms the relative importance of PCC. Moffiet et al. (2005) corroborate this observation, suggesting that canopy openness and spacing contribute greatly to variations in intensity of laser returns due to canopy fractional interception of laser energy within individual footprints.

2. Methods

2.1 Study Site

The study site is The University of Montana's Lubrecht Experimental Forest (LEF) in the Blackfoot River drainage, 54 km northeast of Missoula, Montana, U.S.A. (approximately N 46 ° 53' W 113 ° 27', with elevations from 1160 to 1930 m). LEF covers 11,300 hectares and is dominated by western larch (*Larix occidentalis*) and Douglas-fir (*Pseudotsuga menziesii*) on the north facing slopes and ponderosa pine (*Pinus ponderosa*) on south facing slopes, with a substantial intermixing of species. The eastern portion of the forest is represented by even-aged stands of lodgepole pine (*Pinus contorta*), with subalpine fir (*Abies lasiocarpa*) dominant in higher elevations.

2.2 Field data

Sixty-one rectangular plots (0.04 ha) were collected from five canopy structures, including dense single strata, dense multi strata, moderate single and multi strata, and open strata. The detail characteristics of the strata and plot distribution within these classes are described in Suratno et al. (in review). For the purposes of this study, 43 plots met a dominant species criteria, where dominant species is >70% on a tree basis. The 70% threshold was selected to balance the need for within-plot species homogeneity against number of plots available for analysis. Plots were located using differentially corrected GPS measurements (average accuracy of ± 1.5 m). Of 43 plots, 19 plots were represented by Douglas-fir with an average of 33 trees/plot followed by ponderosa pine (11) consisting of 21 trees/plot, lodgepole pine (6), and western larch (7) as shown in Table 1. The range of tree heights within plots dominated by Douglas-fir (DF) is 8 - 26 m with a mean of ~14 m. Lodgepole pine (LP) plots contain mostly even-aged trees with a mean height of 14 m and mean percent canopy cover (PCC) of 70% . PCC in Table 1 is calculated from laser returns.

Table 1. Plot distribution characteristics of dominant species.

Species	No. Plots	Average trees per plot	Height (m)			Crownbase (m)			PCC (%)		
			Mean	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.
Douglas-fir	19	32.47	14.25	7.96	25.88	7.49	3.22	14.33	58.78	17.45	89.42
Ponderosa pine	11	21.18	16.32	11.43	23.59	8.26	4.15	13.19	40.68	16.84	80.17
Lodgepole pine	6	42.83	13.69	10.11	17.11	6.82	2.69	11.16	70.38	48.87	96.36
western larch	7	36.57	17.15	7.76	24.03	8.97	1.80	11.61	59.12	9.02	89.82

2.3 Lidar Dataset

ALS data were acquired in June 2005 using Leica Geosystems Proprietary (LGP) ALS50. The mean flying height was 1900 meters above ground, mean airspeed was 71 m/s, scan rate was 36 KHz, and scan angle was ± 26 degrees. These parameters resulted in an average return density of 0.44/ m² on the ground, while a laser beam divergence of 1 mrad (1064 nm) combined with flying height resulted in an average footprint size of ~ 1 m². The sidelap was 50% to ensure complete coverage of the study area. A maximum of four returns with intensity for each pulse were recorded depending on vegetation height and pulse return energy following first hits. The vertical accuracy is reported at 0.15 meters and horizontal accuracy is 0.25 meters. Following pre-processing (roll, pitch, and yaw corrections), aboveground laser return points were separated

from ground ‘bare earth’ points using a block-minimum approach in the TerraScan software suite (Terrasolid, Helsinki). A Digital Elevation Model (DEM) was created using Inverse Distance Weighted (IDW) at 1 meter resolution and used to calculate the Canopy Height Model (CHM) using a spot elevation method. The LPS ALS50 used for this study operated with Automatic Gain Control (AGC) to record target intensity, which automatically adjusts the captured raw intensity values for variations in slant range, flying height and system AGC gain (Leica Geosystem, Georgia). This adjustment may have a nonlinear effect on the output target intensity due to reflectance variability. Therefore, intensity correction was carried out to minimize such effect by normalizing the raw intensity with squared range divided by squared average elevation multiplied by AGC value (Korpela et al. 2009).

2.4 Stand Database

The Lubrecht Forest developed a stand database from combinations of serial intensive walk-through inventories and aerial photo interpretations in 1995 and updated it in 2000 (Waterman, 2000). Stand species is defined by a predominant (prime) species and one or more secondary (alpha) species, if the second species composes at least 10% of total canopy cover. Noting that there is ambiguity with the lack of quantitative species distribution data and the fact that some stands are not homogeneous, we assume that stand homogeneity is at least within the dominant species criteria (>70%) if the stand is represented by a prime species and no alpha species. Using this logic, the entire LEF was divided into 1,434 stand polygons, comprised of 180 DF stands (total areas of ~792 ha), 50 PP stands (~257 ha), 102 LP stands (~331 ha) and 12 WL stands (~27 ha). The rest of the stands (1090 polygons) were either “mixed” species or were dominated by subalpine fir (*Abies lasiocarpa*) or Engelmann spruce (*Picea engelmannii*).

2.4 Creating LDA-based species and PCC data (stage 1)

All procedures were performed using regular grids with cell size corresponding to the size of the field plots (20 x 20 m²). Two grid datasets were created: (1) a Linear Discriminant Analysis (LDA)-species layer, and (2) Percent Canopy Cover (PCC). The LDA-species layer was created based on the tree species identification methodology described by Suratno et al. (in review), using the following steps: (1) twelve grid layers were created from the Canopy Height Model, representing means and standard deviations of intensity and canopy height for three return types (all returns, first returns of multiple returns, and single returns (only one return recorded for a given pulse)). Points lower than 2 meters in height were removed to avoid ambiguity between canopy and ground. The variables are listed in Table 2. (2) LDA with cross validation was performed using the previously cited variables on the 43 candidate field plots to classify the four tree species, resulting in 12 coefficients for each of four discriminant functions. For this study, the discriminant function representing Douglas-fir was used to create a single intensity-height interaction layer by applying equation 1, below. The reason Douglas-fir coefficients were used is based on a previous result (Suratno et al., submitted with revision) showing that DF was the most easily distinguished species in the LDA classification. It is worth acknowledging that any of the other three species discriminant functions should produce similar results.

$$\begin{aligned}
 LDA\text{-species} = & 0.97(MIA) + 1.85(SDMIA) - 29.64(MCA) - 32.79(SDMCA) + \\
 & 0.45(MIF) + 0.06(SDMIF) + 18.38(MCF) + 27.76(SDMCF) - \\
 & 0.37(MIS) - 0.47(SDMIS) + 12.55(MCS) + 12.45(SDMCS) - 138.32 \quad (1)
 \end{aligned}$$

The values generated from equation 1 (represented by pixels) were classified into four groups. The thresholds for each group are represented by the LDA mean value from each species

discriminant function plus or minus 10 points (DF: 136.69 ± 10 , PP: 83.98 ± 10 , LP: 74.82 ± 10 , WL: 93.70 ± 10). The 10-point criterion represents approximately one standard deviation. Values that did not meet threshold criteria were identified as a mixed species.

Table 2. Lidar variables derived from intensity, height and return type.

ALS-derived variables	Acronym
Mean intensity of all returns	MIA
Standard deviation of mean intensity of all returns	SDMIA
Mean canopy height of all returns	MCA
Standard deviation of mean height of all returns	SDMCA
Mean intensity of first returns	MIF
Standard deviation of mean intensity of first returns	SDMIF
Mean canopy height of first returns	MCF
Standard deviation of mean height of first returns	SDMCF
Mean intensity of single returns	MIS
Standard deviation of mean intensity of single returns	SDMIS
Mean canopy height single returns	MCS
Standard deviation of mean height of single returns	SDMCS

Additionally, a percent canopy cover (PCC) layer was generated from the lidar point cloud using the proportion of vegetation first returns to total bare earth and vegetation first returns. Both layers were co-registered and combined to generate a single multiband (LDA-PCC) dataset for the final classification.

2.5 Species classification and mapping

Classification and mapping were performed using a supervised Maximum Likelihood Classification (MLC) algorithm. The MLC is based on the probability that a pixel represented by its feature class (A_i with $i \in \{1, 2, \dots, N$ class) located in x,y position of feature vector (B) is assigned into a particular class based on the pixel relative likelihood (probability) occurrence within the probability density function (P) of each class, which is defined by the Bayesian rule (Jensen and Nielsen, 2007) as:

$$P(A_i | B_{x,y}) = \frac{P(B_{x,y} | A_i)P(A_i)}{P(B_{x,y})} \quad (2)$$

where $P(A_i/B_{x,y})$ is the probability or likelihood that the hypothesis is true for feature class A given feature vector B , $P(B_{x,y}/A_i)$ is the conditional probability of feature vector B given feature class A , $P(A_i)$ and $P(B_{x,y})$ are the prior probabilities of the class A and pixels represented by the feature vector $B_{x,y}$ respectively.

The MLC assumes that species class distributions on both LDA-species and PCC layers are normally distributed and we tacitly accept this assumption while acknowledging that it is difficult to assess with the small number of plots represented by each species. Plots were used as training areas for each species class and the MLC was completed with equal prior probabilities for each class. The species classification was applied to the entire 11,300 ha study area. Accuracy assessments were performed on a per pixel basis for the four species by calculating the proportion of pixels correctly classified (by species) to the total number of pixels in the analysis. Pixels classified as mixed were not included in the accuracy assessment, but are discussed below.

3. Results and Discussion

Figure 1 shows a subset of the study area, including: (a) stand-based species data, (b) PCC, (c) LDA-species, and (d) MLC-species. Application of the LDA function results in a grainy distribution of species that falsely emphasizes western larch and mixed classes. MLC resolves much of this ambiguity while improving the discrimination of LP and PP. We note that WL is not apparent in the MLC classification and instead, grouped with PP and DF. The fractions of mix-species also decrease from LDA to MLC. Regarding the mixed species stands, we observe that most of the landscape belongs to this class. Although difficult to depict in the stand species layer (Figure 1a), the lidar-base prime species in each mixed polygon matches the prime species of adjacent single species polygons and the prime species in the stand database. However, we do not currently have the field data to support validation in mixed species stands.

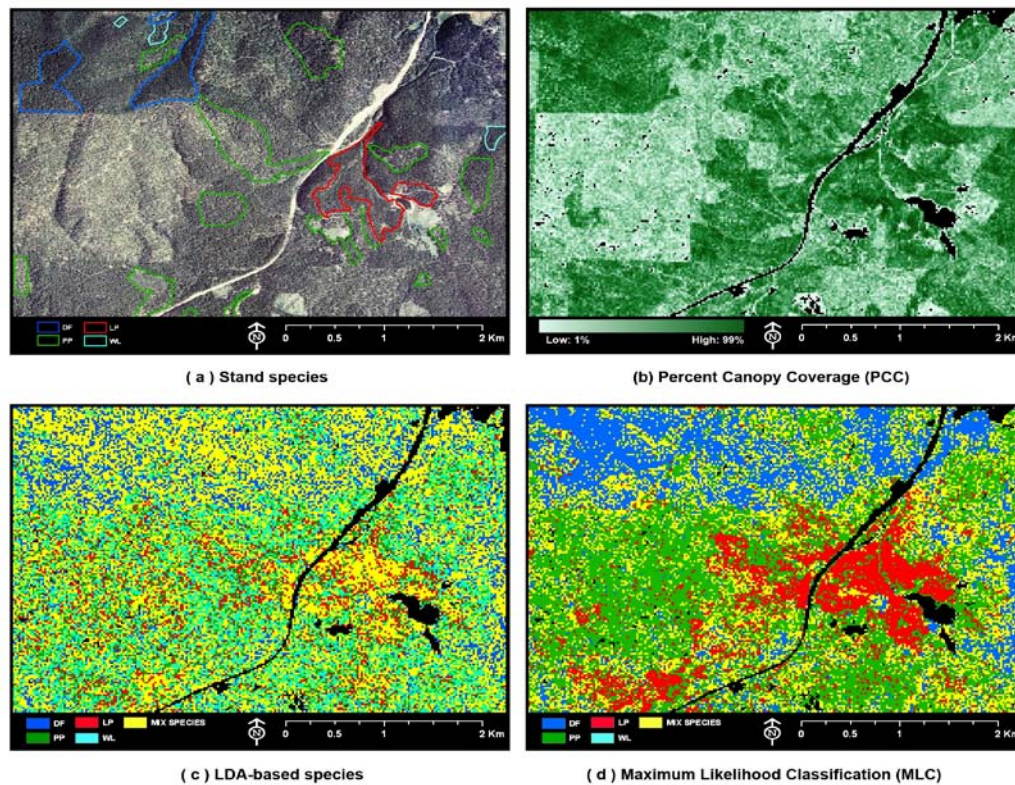


Figure 1. A subset of classification layers within study area.

The error matrix for species classification is shown in Table 3. The classification using LDA method produces overall accuracy of 45% while the classification using MLC method generates overall accuracy of 75%. By using LDA, the highest producer's accuracy (the total of correctly classified pixels from the class samples) is DF (75%) followed by WL, LP, and PP (39%, 21%, and 18% respectively). Meanwhile, the user's accuracies (the correctly assigned pixels to one specific class) of DF, LP, PP, and WL are 79%, 65%, and 3% respectively. On the other hand, by using MLC, the highest producer's accuracy is DF (86%) followed by LP and PP (68% and 57% respectively). However, class WL produces an accuracy of 0%. Meanwhile, the user's accuracies of DF, LP, and PP are 84%, 75%, and 49%. Identical to the producer's accuracy, class WL generates an accuracy of 0%.

Table 3. Error matrix for species classification

LDA Classification						MLC Classification					
		Field reference						Field reference			
A. Species	DF	PP	LP	WL	Row total	A. Species	DF	PP	LP	WL	Row total
DF	8037	1220	728	167	10152	DF	17072	1859	1000	298	20229
PP	506	931	1899	110	3446	PP	2055	3646	1656	104	7461
LP	222	575	1607	81	2485	LP	668	931	5612	291	7502
WL	1989	2364	3451	218	8022	WL	1	1	1	0	3
Column total	10754	5090	7685	576	24105	Column total	19796	6437	8269	693	35195
B. Classification accuracy						B. Classification accuracy					
Producer's accuracy			User's accuracy			Producer's accuracy			User's accuracy		
DF = 74.73%			DF = 79.17%			DF = 86.24%			DF = 84.39%		
PP = 18.29%			PP = 27.02%			PP = 56.64%			PP = 48.87%		
LP = 20.91%			LP = 64.67%			LP = 67.87%			LP = 74.81%		
WL = 37.85%			WL = 2.72%			WL = 0.00%			WL = 0.00%		
Overall accuracy = 44.77%						Overall accuracy = 74.81%					

The results above suggest that it may be possible to use relatively low density airborne laser scanning data to classify and map trees in mixed coniferous forests like those found at LEF. However, it is noteworthy that classification accuracy is low when mixed species are considered and there is still uncertainty about how well the approach captures stand heterogeneity. Almost without exception, the classification identifies the correct mix of species within each mixed polygon. Yet, the mixed validation fails because the classification produces few mixed pixels relative to species-specific pixels. Therefore, during the validation, we are comparing species pixels (classified) to mixed pixels (truth) and can only conclude that we misclassify most of the time. Validation at stand-level instead of pixel level would perhaps begin to illuminate the effectiveness of the approach in heterogeneous stands.

The higher classification accuracy for DF is probably due to the larger number of samples distributed across diverse stands, and the low accuracy for WL is conversely attributable sample-size constrained misrepresentation of species variability. Because the MLC uses probability to apportion pixels into classes, a representative probability density function for each attribute (per species) is essential. Additionally, WL almost always occur secondary to PP and DF (Arno, et al. 1985). The high classification accuracy of LP is likely due to the fact that laser return intensity is generally lower than for other species (Suratno et al., submitted with revision) and the species usually occurs in even-aged stands.

As previously noted, percent canopy cover strongly affects return intensity due to fractional interception of energy by scattered foliage within the laser footprint. Consequently, the addition of PCC to the classification is logical and improves the result. We note that PCC was included as a variable in the LDA (results not presented here) with little accuracy improvement to the LDA classification. We speculate that the reason PCC improves the MLC classification but not the LDA classification is because the interaction between intensity and PCC is non-linear. The MLC is not constrained in the same way and instead relies on probabilities.

4. Conclusions

The research presented here represents a first attempt at classification with promising results and suggests that ALS-data can be used to classify and map four western North American conifer species. The Maximum Likelihood Classification produced an overall accuracy of 75% with the best results for Douglas-fir followed by lodgepole pine and ponderosa pine, while western larch was difficult to identify. The ambiguity of species distributions in the stand database used for ground truth in mixed forest prevented robust validation for a large fraction of

the landscape. However, the classification resulted in the correct species appearing in each stand polygon and a stand based validation using the prime/alpha notation is warranted. The inclusion of prior probability weights informed by biophysical attributes such as slope, aspect, and elevation will likely clear up some of the uncertainty between species that exhibit strong site preferences. Additionally, it may be possible to improve the classification by increasing the number of training samples to better represent variability. We note that the samples were selected based on structure type rather than species. Lastly, we highlight the simplicity and repeatability of methods presented, but acknowledge considerable uncertainty regarding classification performance in mixed species stands.

Acknowledgements

This study was funded by the National Center for Landscape Fire Analysis (NCLFA) in the College of Forestry and Conservation of the University of Montana and The USDA Forest Service McIntire-Stennis Program. We thank Eric Rowell, Crystal Stonesifer, Casey Teske, Erik Hakanson, Tim Wallace, Ann Hadlow, Josh Rodriguez, Martin Twer, and R.J. Hannah for field assistance. We also would like to thank Ron Roth (Leica Geosystems, Inc.) for his communications on technical aspects of ALS50 System.

5. References:

- Arno, S.F., D.G. Simmerman, and R.E. Keane, 1985. Forest succession on four habitat types in Western Montana. General Technical Report INT-177. USDA Forest Service, Intermountain Forest and Range Experiment Station Ogden, Utah.
- Brandtberg, T., T.A. Warner, R.E. Landenberger, and J.B. McGraw, 2003. Detection and analysis Of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment*, 85, 290–303.
- Brandtberg, T., 2007. Classifying individual tree species under leaf-off and leaf-on conditions Using airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5): 325-340.
- Donoghue, D.N.M., P.J. Watt, N.J. Cox, and J. Wilson, 2007. Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. *Remote Sensing of Environment*, 110, 509-522.
- Guohui, H., M.D. Desai, and X. Zhang, 1999. Minimum component eigen-vector based classification technique with application to TM images. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 15-19 March 1999. Phoenix, Arizona.
- Holmgren, J., and A. Persson, 2004. Identifying species of individual trees using airborne laser scanning. *Remote Sensing of Environment*, 90, 415-423.
- Jensen, F.V., and T.D. Nielsen, 2007. *Bayesian Networks and Decision Graphs*. Second Ed. Springer. New York. 447 p.
- Jia X., and J.A. Richards, 1994. Efficient maximum likelihood classification for imaging spectrometer datasets. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2): 274-281.
- Jia X., and J.A. Richards, 2003. Efficient transmission and classification of hyperspectral image data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(5): 1129-1131.
- Jieping Ye, T. Xiong, and R. Janardan, 2006. CPM: A Covariance-preserving Projection Method. *Proceedings: Siang Conference on Data Mining 2006*. Bethesda, Maryland, April 20-22, 2006.
- Korpela, I.S., 2008. Mapping of understory lichens with airborne discrete-return LiDAR data. *Remote Sensing of Environment*, 112, 3891–3897.
- Krzanowski, W.J., 1988. *Principles of Multivariate Analysis: A User's Perspective*. Clarendon Press. Oxford. 563 pp.
- Lillesand, T.M. and R.W. Kiefer, 2000. *Remote Sensing and Image Interpretation* 4th Edition, John Wiley & Sons, New York 724 p.

- Moffiet, T., K. Mengersen, C. Witte, R. King, and R. Denham, 2005. Airborne laser scanning: Exploratory data analysis indicates potential variables for classification of individual trees of forest stands according to species. *ISPRS Journal of Photogrammetry and Remote Sensing*. 59, 289-309.
- Orka, H.O., E. Naesset, , and O.M. Bollandsas, 2007. Utilizing airborne laser intensity for tree species classification. *ISPRS Workshop on Laser Scanning and SilviLaser 2007*, Espoo, Finland, September 12-14, 2007.
- Persson, A., Holmgren, J., and Söderman, U., 2002. Detecting and Measuring Individual Trees Using an Airborne Laser Scanner. *Photogrammetric Engineering and Remote Sensing*, 68(9): 925-932.
- Suratno, A., C. Seielstad, L. Queen, 2009. Tree Species Identification in Mixed-Coniferous Forest Using Airborne Laser Scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*. Submitted with revision.
- Waterman, J.R., 2000. Building a Lubrecht forest inventory internet mapping service and volume comparisons. The Univeristy of Montana. Missoula, MT.