UTILIZING AIRBORNE LASER INTENSITY FOR TREE SPECIES CLASSIFICATION

H. O. Ørka, E. Næsset, O. M. Bollandsås

Dept. of Ecology and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway – (hans-ole.orka, erik.naesset, ole.martin.bollandsas)@umb.no

KEY WORDS: Laser scanning, High resolution, Forest, Inventory, Analysis, Classification

ABSTRACT:

High-resolution datasets from Airborne Laser Scanning (ALS) provide information to extract the outline of single tree crowns. Laser echoes with spatial coordinates inside these single-tree crowns give the ability of measuring biophysical properties and to classify species of these single-trees. Species classification by ALS-data is based on differences in crown shape, crown density, reflectivity and distribution of foliage and branches between tree species. All of these parameters may be expressed by spatial coordinates of the point cloud or by the intensity of the backscattered signal measured by ALS. In this study we investigate mean intensity and standard deviation of intensity computed for single trees by explorative data analysis and linear discriminant analysis. We explore differences in spruce, birch, and aspen trees for different echo categories from a multiple return ALS system. We found that intensity could assist species discrimination. The overall classification accuracies obtained were from 68 to 74 %, depending on number of variables considered. In spite of the heterogeneous structure of the forest studied, the classification accuracy was fairly high. Intensity metrics computed from different echo categories influence overall accuracies by 3 to 4 %, depending on the intensity metric used in the classification. Both species reflectivity and structural characteristics within the tree crown will influence intensity recorded by ALS.

1. INTRODUCTION

In the past two decades, the capabilities of Airborne Laser Scanning (ALS) in forest structural assessment have been documented in a many research studies as well as in operational inventories. Today, ALS-sensors provide high-resolution datasets with 5-10 points per square meter. From these high-density datasets single trees can be extracted and measured (e.g. Persson et al. 2002; Morsdorf et al. 2004; Solberg et al. 2006). The distribution of laser echoes in a tree is a result of crown shape, crown density, distribution of foliage mass and branches, surface reflectivity, etc. Since many of these parameters may represent distinct characteristic of different tree species, ALS have been proposed and tested to support single-tree species classification. In addition to the distribution of the spatial coordinates x, y, and z for the returned echoes, most proprietary airborne lasers are also capable of measuring the reflectivity of the surface hit by the laser pulses. The reflectivity measured by pulse lasers represents the intensity of the maximum energy of the returned echo (Wehr and Lohr 1999). This intensity recorded by ALS could also assist tree species classification if species are separable in the spectral domain at the wavelength emitted by the laser (e.g. 1064 nm). Already in 1985, Schreier et al. (1985) found that the mean reflection and reflection variability measured by a airborne laser system could be used to differentiate between coniferous and deciduous trees. Later, intensity metrics derived from ALS have been used in tree species classification studies in boreal conifer forest in Scandinavia (Holmgren and Persson 2004), North American deciduous forest (Brandtberg et al. 2003; Brandtberg 2007), and sub-tropical Australian forest (Moffiet et al. 2005). In some of these classification studies the echo categories, i.e., first and last echoes, are treated separately, in other studies echo categories are joined together. However, combining all echo categories are not recommended because separate echo categories may play an important role in achieving better total classification accuracies (Brandtberg 2007). As use of ALS systems capable of record multiple echoes become more common, knowledge of intensity behavior of different echo categories will be important in further species classification studies. The objectives of the present study were to characterize and analyze ALS-derived intensity metrics for (1) different tree species and 2) for different echo categories, and 3) to test and report classification performance of ALS-derived intensity metrics in a heterogeneous boreal forest.

2. MATERIAL

2.1 Study area

The study area is located in the Østmarka natural reserve, in southeastern Norway (59°50'N, 11°02'E, 190-370 m a.s.l). The area has not been logged or managed over the last 60 years. Today the forest appears with large within-stand variation of tree ages and tree sizes. The dominating species are Norway spruce (*Picea abies* L.) and Scots Pine (*Pinus silvestris* L.). In addition, deciduous species, mainly Birch (*Betula ssp.*) and Aspen (*Populus tremula* L.), are found scattered in the landscape.

2.2 Field data

The field inventory was carried out during summer 2003 on 20 sample plots of size 0.1 ha. The sample plots were located in the spruce-dominated areas. Polar coordinates relative to plot center of all trees with diameter at breast height greater than 3 cm were registered. Plot center coordinates were determined by differential Global Navigation Satellite Systems (GNSS). On each plot, sample trees were selected according to two different sampling schemes. First, eight trees were selected by choosing the first dominant tree in each cardinal direction and the nearest tree to each of the four dominant ones. Second, all deciduous trees on the plot and some deciduous trees just outside the plot were selected. This was done to increase the number of deciduous trees in the sample. On sample trees height, diameter in breast height, and crown radius were measured. Crown radius was determined as the mean of the radii measured in the four cardinal directions. Further details about the field work and the georeferencing procedures can be found in Solberg et al. (2006) and Bollandsås and Næsset (2007).

2.3 Airborne laser scanner data

ALS-data used in this study were acquired 18 June 2005 with the Optech ALTM 3100 sensor. Table 1 provides information about laser scanner settings planned for the acquisition. The initial processing of laser data was accomplished by the contractor (Blom Geomatics, Norway). First, all echoes outside a half scan angle of 8 degrees were deleted. Ground points were classified using the Terrascan software (Anon. 2004) and the ground-classified points were used to build a Triangular Irregular Network (TIN) terrain model. The TIN heights were then subtracted for each laser echo to produce height above the ground surface. Furthermore, all echoes with height less than 1.3 m where deleted from the dataset. In previous studies, this Ground Threshold Value (GTV) has typically been set to 2 m (Nilsson 1996; Næsset 2002, 2004). The choice of 1.3 m as GTV is based on the requirement of consistency with other measurements on the trees. For example, stem diameter and age are usually measured at breast height (1.3 m). In this multi-layered forest, we also wanted to keep as much information about small trees as possible and for that reason we did not want to exclude tall under-vegetation.

Parameters	
Flying altitude AGL (m)	750
Pulse repetition frequency (Hz)	100000
Scanner frequency (Hz)	70
Half scan angle (deg.)	10
Flying speed (m/s)	75
Swath (m)	264
pt/m^2	5.09
Beam divergence (mrad)	0.26

Table 1. Planned laser scanner settings for data acquisition

The elapsed time between field measurements and ALS data acquisition represented two years of growth. The forest in the nature reserve is in general old and little growth is seen. Hence, the field measurements were not adjusted for growth.

The Optech ALTM 3100 sensor is a multiple return system capable of recording up to four echoes per pulse. Older systems (e.g. Optech ALTM 1233) record two echoes for each pulse, i.e., first and last echoes. The number of echoes recorded by the ALTM 3100 depends on the returned energy of the laser pulses and the vertical resolution, i.e., the minimum distance between separate echoes. For the ALTM 3100, the vertical resolution is 2.1 m between the two first returns and 3.8 m for the other returns. If four returns are detected the sensor label them as “first echo of many”, “second echo”, “third echo” and “last echo”. If only two echoes are identified, these are labeled “first echo of many” and “last echo”. If three echoes are recorded, these are labeled “first echo of many”, “second echo” and “last echo”. If only one significant return is identified, this is labeled as an “only echo”. In this study, the ALS data were delivered by the contractor to be as close as possible to the structure of data provided by the old ALTM 1233 sensor. This was done by combining “first echoes of many” and “only echoes” in one dataset and “last echoes” and “only echoes” in another dataset. The reason for this was to be able to compare the two different sensors, i.e., the ALTM 1233 and 3100 instruments. However, in this particular study we split the dataset into the original echo categories, i.e., “first echoes of many”, “last echoes”, and “only echoes”. We did this by extracting and labeling echoes with the same spatial coordinates in the two datasets as “only echoes” and deleting them from the two original datasets. These three echo categories were used in our analysis and for short they are denoted “first”, “only” and “last” echoes.

3. METHODS

In this study, no single-tree detection based on the ALS-data was carried out. Instead we buffered the field-derived tree stem positions with the mean crown radius. Laser echoes inside these circular crown segments were then tied to the buffered tree. The forest in the natural reserve is partly multi-layered and in case of overlapping tree crowns echoes were tied to the tallest tree.

For each tree we computed the mean intensity (MI) and the standard deviation of intensity (SDI) based on the laser point cloud. These two metrics are selected for this study because thet are the most frequently used in other single-tree species classification studies (Schreier et al. 1985; Brandtberg et al. 2003; Holmgren and Persson 2004; Moffiet et al. 2005; Brandtberg 2007). These two intensity metrics were computed for each echo category. In total, six variables where therefore derived. In this paper, the last character of the variable name represents the echo category, i.e., F=first, O=only and L=last. For example, mean intensity of first echoes of many are denoted MIF, mean intensity of only echoes MIO, etc. To be able to calculate the SDI at least two echoes are needed. Thus, in order to have a consistent dataset without missing values, all trees hit by less than two echoes in an echo category were discarded from further analysis. In addition, only Norway spruce (S), birch (B), and aspen (A) trees were considered.

Explorative data analysis was performed using graphical methods by means of box-and-whisker plots (Tukey 1977; Anon. 2006). The box-and-whisker plot will give an overview of the data showing first and third quartile as the box (“hinges”), the median as the horizontal line dividing the box and extreme values as points outside the “whisker” defined as:

$$+/-1.58\frac{IQR}{\sqrt{n}} \tag{1}$$

where n = number of observations
IQR = Inter Quartile Range (i.e. difference between first and third quartiles)

In addition to this graphical representation of summary statistics, we explored relationships between intensity metrics by computing Pearson correlation coefficients. We also carried out a Principal Component Analysis (PCA) to explore and visualize variance in all computed metrics. The PCA calculation was done by a singular value decomposition of the centered data matrix.

The investigation of classification performance was conducted with Linear Discriminant Analysis (LDA). LDA was carried out using equal prior probabilities and leave-one-out cross validation by *lda* function of the R-package MASS (Venables and Ripley 2002). Classification was tested for all variables on independent basis. Overall accuracy was calculated as the percent trees correctly classified over total number of trees in the analysis. Then two to six (all) variables were combined and the combination(s) with the highest overall accuracy were selected. These overall accuracies are reported in Table 4. An error matrix for the combination of variables that produced the highest overall accuracy is also reported (Table 5).

4. RESULTS

Of the 260 field measured trees a total of 224 trees were of the tree species considered and hit by at least three echoes of each category and included in the analysis. Summary statistics of trees used in analysis appear in Table 2.

	Spruce	Birch	Aspen
Number of trees	133	70	21
Tree height (m)	23.1 (4.7)	18.6 (4.2)	23.8 (3.9)
Stem diameter (cm)	31.7 (8.3)	22.8 (7.5)	33.4 (8.3)
Crown radius (m)	1,78 (0.46)	2.02 (0.56)	2.51 (0.81)

Table 2. Mean values and standard deviations (in parenthesis) of field measured trees used the in analysis.

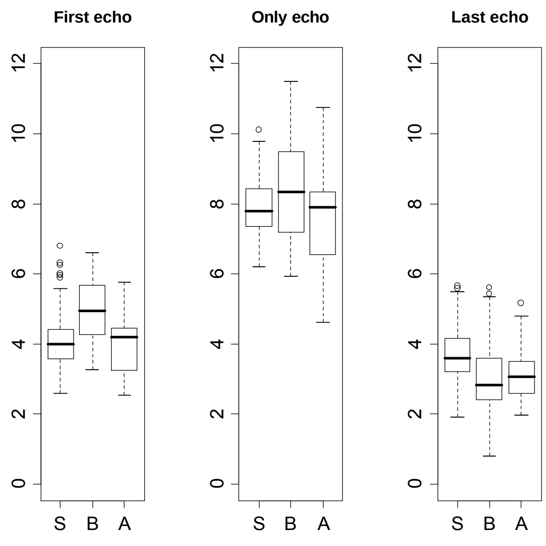

Figure 1. Mean intensity for different echo categories and tree species

Explorative data analysis by means of box-and-whisker plot of MI and SDI provides an overview of the variables (Figures 1 and 2).

The correlation matrix of intensity-derived metrics (Table 3) unveiled two relationships, i.e., (1) a positive correlation between MI and SDI in each echo category and (2) a positive correlation between metrics computed from first echoes and only echoes. Correlations between first and only echoes are found between MI and SDI computed from these echoes and between MIO and SDIF. All correlations in the matrix are stronger for birch and aspen compared to spruce when correlation analyses were preformed on subsets containing separate tree species.

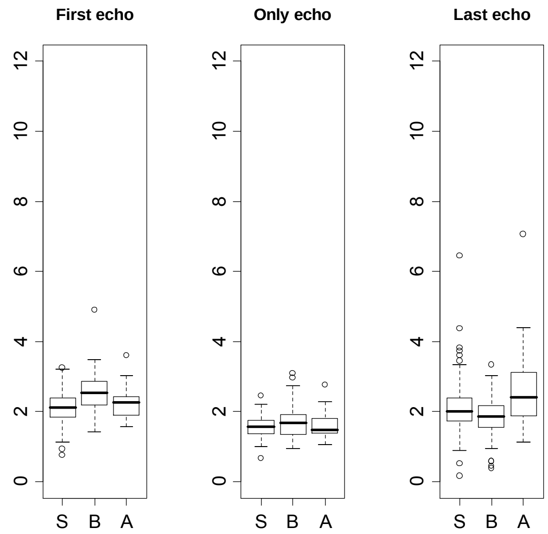

Figure 2. Standard deviation of intensity for different echo categories and tree species

	MIF	MIO	MIS	SDIF	SDIO	SDIS
MIF	1.00					
MIO	**0.57**	1.00				
MIS	-0.17	0.13	1.00			
SDIF	**0.44**	**0.59**	0.02	1.00		
SDIO	0.29	**0.47**	0.06	**0.45**	1.00	
SDIS	-0.08	0.01	**0.42**	0.11	0.14	1.00

Table 3. Correlations between intensity metrics

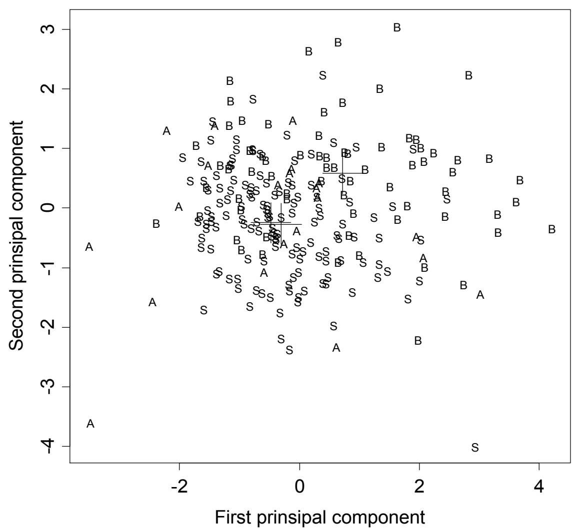

Figure 3. PCA score plot of first and second principal component. Class centers indicated by crosses.

The two first components of the PCA explained 73 % of the variation in intensity-derived metrics whereas the first component explained 46 % only. The score plot of the two first components is displayed in Figure 3 labeled with tree species.

LDA classification was performed both by using only one variable at a time and for combinations of variables (Table 4). The best classification obtained included mean intensity for first and last echoes and standard deviation of intensity for all echo categories. The error matrix (Table 5) for this classification provided an overall accuracy of 74.1 % and a *Kappa* coefficient of 0.49. The classification accuracies for the single species were 87.2 % for spruce, 64.3 % for birch and 23.8 for aspen.

Combination of intensity metrics	Overall accuracy (%)
MIF	**68.3**
MIO	65.2
MIS	66.5
SDIF	64.3
SDIO	62.5
SDIS	60.3
MIF, MIS	71.4
MIF, MIS, SDIF	72.8
MIF, MIS, SDIO	72.8
MIF, MIS, SDIF, SDIS	73.7
MIF, MIS, SDIF, SDIO	73.7
MIF, MIS, SDIF , SDIO, SDIS	**74.1**
All intensity metrics	73.2

Table 4. Overall accuracy of species classification for different combinations of laser intensity metrics

	Field reference		
Classification	Spruce	Birch	Aspen
Spruce	**116**	25	12
Birch	13	**45**	4
Aspen	4	0	**5**

Table 5. Error matrix for tree species classification

5. DISCUSSION

The distribution of trees on different species and size classes in the natural reserve is different from what we will find in a managed forest. The structure of the forest limits the number of deciduous trees. The birch trees in this study are also somewhat smaller than the spruce trees and the aspen trees are in general large and old with tick branches that dominates the crown. However, our goal was to get some basic experience with intensity metrics for three species discrimination and for this purpose we found the data relevant.

In the box-and-whisker plots, the most pronounced pattern is the higher values of mean intensity of only echoes (Figure 1). The mean intensities of only echoes are about the size of the sum of first and last echoes. This is reasonable size only echoes are recorded when there is not enough energy to produce a second return (or vertical distance is too short). However, both first and only echoes should be returned from the same canopy surface. The difference in intensity from these echoes indicates that they have different origins not only depending on species reflectivity. Thus, it seems that laser intensity is a result of other tree characteristics than the species-specific reflectivity (Moffiet et al. 2005). It is therefore likely that only echoes are returned from denser parts of the tree like the stem or thick branches. However, the observed differences between echo categories are important and indicate that treating different echoes separately in species classification may provide more information than combining them.

The box-and-whisker plots (Figures 1 and 2) indicate that the differences between tree species are relatively small. The median values of each species are in general between the first and third quartile of the other species they are compared to. The exception is first echoes which are different in both MI and SDI for spruce compared to birch. We expected that deciduous species would have nearly the same reflection in the wavelength emitted by the laser. Surprisingly, intensity metrics from first and only echoes of aspen trees were more similar to spruce than birch. This similarity may be explained by reflectivity of bark (branches and stems) and by crown characteristics such as density and structure in addition to foliage reflectivity. Last echoes of spruce have higher mean intensity than the two deciduous species. A likely reason is the relation to first echoes where spruce trees have more energy left to produce a second return. However, the correlations between metrics derived from first and last echoes are weak (Table 3).

Correlations between mean intensities and standard deviations within echo categories at plot level are also reported by Moffiet et al. (2005). They explain the correlation between mean intensity and standard deviation of intensity by the interaction of the features of forest and trees on the footprint. A permeable surface will return a small portion of the pulse energy and it will need more time before a significant return will be recorded. This will lower the mean intensity and increase the standard deviation of intensity compare to a denser surface. However, we found that these relationships are stronger in deciduous species. Hence, the effect is different between tree species. The correlations between metrics from first echoes and only echoes are also interesting, especially as first echoes seem to be more important in separating species than only echoes. The correlations between metrics of first echoes and only echoes indicate effects which influence both echo categories in the same way. However, first echoes are better in discriminate species. Hence, first echoes are probably less affected by these overall effects.

The explorative data analysis and the principal component analysis indicated large variation in the intensity metrics. Only parts of the variation were explained by the first two components of the PCA, indicating that much of the variation in these metrics will be difficult to explain. A factor of variance that may be used in further studies is the incident angle of the laser pulse. However, as shown in Figure 5, some of the variation is explained by tree species. The birch and spruce trees are nearly separated into two groups in the PCA score plot. This separation of tree species in the score plot supports the use of laser intensity in discriminating between these two species. The score values for aspen trees are found more scattered and mixed with score values of spruce trees in the plot. The scattered score values of aspen trees may be an effect of the special crown structure or by the limited number of observations in the dataset. However, the observation of different score values in aspen and birch trees indicate that these tree species should be treated separately and not as a common group. In addition, both the box-and-whisker plots and the score plot point out that the variance in intensity metrics, e.g. MIO, is smaller for spruce than for deciduous species. This may be due to the fact that spruce trees have a more uniform crown than deciduous trees.

The study site is located in a heterogeneous forest with a number of suppressed trees. They are indeed difficult to detect and segment by automated single-tree segmentation as well (Solberg et al. 2006), and these trees may also be the most difficult ones to classify. In spite of the complex stand structure, the classification accuracy was fairly high. The cross validated classification indicated that intensity of first echoes are the most appropriate ones to separate spruce, birch, and aspen trees. In addition, intensity metrics from last echoes and the standard deviation of intensity from only echoes will contribute with information separating the three species. The classification is explorative and more appropriate training- and testing datasets are required to verify these initial findings.

6. CONCLUSION

The results of this study indicate that intensity from high-resolution airborne laser could assist species discrimination. Of the tested echo categories the ”first echoes of many” provided most information for discriminating between species. Last echoes provided additional information and ”only echoes” provided least information for separating species. Intensity measured by ALS is also highly variable. The high intensity of ”only echoes” compared to first echoes and relationships between echo categories indicated that other tree structural characteristics than species reflectivity are of importance for the intensity value. The classification accuracies obtain were form 68 to 74 %, depending on number of variables considered. Using additional ALS-derived metrics, like canopy density- or height metrics, will probably improve classification performance. Hence, additional metrics and information will be use in and operational application. In the near future, use of integrated systems like sensors combining laser and image sensors are likely to become more common. Data from such systems will have a great advantage in tree species classification (Persson et al. 2004).

7. REFERENCES

Anon. (2004). TerraScan user's guide. Terrasolid, Helsinki. http://terrasolid.fi

Anon. (2006). R: A Language and Environment for Statistical Computing (R Development Core Team). R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Bollandsås, O.M., & Næsset, E. (2007). Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data *Scandinavian Journal of Forest Research,* 22(1), pp. 33-48.

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), pp. 325-340.

Brandtberg, T., Warner, T.A., Landenberger, R.E., & McGraw, J.B. (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(3), pp. 290-303.

Holmgren, J., & Persson, Å. (2004). Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment,* 90(4), pp. 415-423.

Moffiet, T., Mengersen, K., Witte, C., King, R., & Denham, R. (2005). Airborne laser scanning: Exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. *ISPRS Journal of Photogrammetry and Remote Sensing,* 59(5), pp. 289-309.

Morsdorf, F., Meier, E., Kotz, B., Itten, K.I., Dobbertin, M., & Allgower, B. (2004). LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment,* 92(3), pp. 353-362.

Nilsson, M. (1996). Estimation of tree weights and stand volume using an airborne lidar system. *Remote Sensing of Environment,* 56(1), pp. 1-7.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment,* 80(1), pp. 88-99.

Næsset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research,* 19(2), pp. 164-179.

Persson, Å., Holmgren, J., Soderman, U., & Olsson, H. (2004). Tree species classification of individual trees in Sweden by combining high resolution laser data with high resolution near-infrared digitalt images. In: M. Thies, B. Koch, H. Spiecker & H. Weinacker (Eds.), *Laser-Scanners for Forest and Landscape Assessment.* Freiburg, Germany.

Persson, Å., Holmgren, J., & Söderman, U. (2002). Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing,* 68(9), pp. 925-932.

Schreier, H., Lougheed, J., Tucker, C., & Leckie, D. (1985). Automated measurements of terrain reflection and height variations using an airborne infrared laser system. *International Journal of Remote Sensing,* 6(1), pp. 101-113.

Solberg, S., Næsset, E., & Bollandsås, O.M. (2006). Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogrammetric Engineering and Remote Sensing,* 72(12), pp. 1369-1378.

Tukey, J.W. (1977). *Exploratory data analysis*. Addison-Wesley Reading, pp. 688

Venables, W.N., & Ripley, B.D. (2002). *Modern Applied Statistics with S.* Springer, New York, pp. 495

Wehr, A., & Lohr, U. (1999). Airborne laser scanning - an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing,* 54(2-3), pp. 68-82.