The Effect of Seasonal Variation on Automated Land Cover Mapping from Multispectral Airborne Laser Scanning Data

Kirsi Karila[a*], Leena Matikainen[a], Paula Litkey[a], Juha Hyyppä[a], Eetu Puttonen[a]

[a]*Finnish Geospatial Research Institute at the National Land Survey of Finland, Centre of Excellence in Laser Scanning Research, FI-02430 Masala, Finland*

*corresponding author email: kirsi.karila@nls.fi

Abstract

Multispectral airborne laser scanning (MS-ALS) sensors are a new promising source of data for automated mapping methods. Finding an optimal time for data acquisition is important in all mapping applications based on remotely sensed datasets. In this study, three MS-ALS datasets acquired at different times of the growing season were compared for automated land cover mapping and road detection in a suburban area. In addition, changes in the intensity were studied. An object-based random forest classification was carried out using reference points. The overall accuracy of the land cover classification was 93.9% (May dataset), 96.4% (June) and 95.9% (August). The use of the May dataset acquired under leafless conditions resulted in more complete roads than the other datasets acquired when trees were in leaf. It was concluded that all datasets used in the study are applicable for suburban land cover mapping, however small differences in accuracies between land cover classes exist.

Keywords: multitemporal, multispectral, ALS, lidar, land cover, classification, roads, laser scanning

MANUSCRIPT

1. Introduction

Recently, automated mapping and change detection methods have become feasible for urban mapping due to new high-resolution remotely sensed datasets. Airborne laser scanning (ALS) is a technique widely used in digital elevation model (DEM) production and other 3D mapping applications. Nowadays, in addition to the XYZ coordinates, the intensity of the returning pulse is also recorded. In recent studies, ALS intensity has been found suitable for

many applications, including urban land cover classification (Hug and Wehr 1997; Guo et al. 2011; Zhou 2013; Yan, Shaker, and El-Ashmawy 2015). However, the class separation using single wavelength intensity data is limited. Therefore, numerous studies have combined ALS height/geometry information with multispectral optical data, often resulting in enhanced classification accuracy (e.g. Gamba and Houshmand 2002; Huang et al. 2008; Salah, Trinder, and Shaker 2009; Matikainen and Karila 2011).

A multispectral ALS (MS-ALS) sensor simultaneously provides, in addition to accurate height information, the intensity of reflected laser pulses at more than one wavelength. In 2014, the first commercial MS-ALS sensor was launched by Teledyne Optech (Toronto, Ontario, Canada). Recent studies have shown that MS-ALS is suitable for land cover classification (Wichmann et al. 2015; Fernandez-Diaz et al. 2016; Bakuła, Kupidura, and Jełowicki 2016; Matikainen et al. 2017; Morsy, Shaker, and El-Rabbany 2017; Teo and Wu 2017), forest and vegetation mapping (Hopkinson et al. 2016; Nabucet et al. 2016; Leigh and Magruder 2016; Yu et al. 2017; Budei et al. 2018) and road mapping (Karila et al. 2017). In comparison to optical aerial images, the independence of illumination conditions and the lack of shadows are real advantages of the data. These properties also mean that MS-ALS data have great potential for further increasing the automation level in mapping (Matikainen et al. 2017).

The accuracy of mapping is strongly dependent on the quality of the input data. In traditional mapping based on aerial images, the acquisition time is selected based on the mapping application. This type of mapping has normally been carried out by human operators who interpret the image data visually. For the built-up land cover mapping, spring images under leafless conditions are often preferred since there is not so much occlusion of man-made objects by the vegetation. For vegetation mapping (species identification), summer (trees in leaf) or fall images are preferred. In addition, in several studies multi-seasonal data

MANUSCRIPT

have been found to be better than single season data for automatic vegetation mapping (e.g. Heinl and Tappeiner 2012; Clark 2017). Automatic methods are often more effective than a human operator at combining information from multiple sensors, bands or acquisition times.

It can be expected that automated analysis methods can have different preferences for data acquisition time than visual methods. For example, the mapping of roads is easier for a human operator in leaf-off rather than leaf-on conditions. The tree cover can also be a problem for automatic classification algorithms. Generally, however, summer conditions with green vegetation might help automatic classification algorithms to separate roads from the surrounding environment. Only a few studies on the effect of seasonal variation on the accuracy of land cover classification exist, with tree species classification being the most studied topic. A multitemporal comparison of tree species classification has been carried out using hyperspectral imagery and lidar (Voss and Sugumaran 2008) as well as optical satellite data (Key et al. 2001; Sugumaran, Pavuluri, and Zerr 2003). There is thus a need for further studies on the topic in order to understand the effect of season on the performance of automated mapping methods and select the most optimal data acquisition times for different purposes.

Another important question related to multitemporal MS-ALS data is the stability of the intensity values for objects that do not change during the year. It affects the feasibility of the intensity data from two separate data acquisitions for direct automatic change detection. The stability of the intensity of operational MS-ALS acquisitions and how seasonal variation affects the intensity of different land cover classes have not been reported yet.

Generally, the recorded laser intensity is affected by the variations in range, incidence angle, emitted power, atmosphere and noise. It has been reported that ALS intensity levels between different flights vary (Kaasalainen et al. 2011). Studies have also shown that radiometric pre-processing (correction) of ALS data is beneficial (Höfle and Pfeifer 2007;

MANUSCRIPT

Yan and Shaker 2014; Yan and Shaker 2017), and that for absolute calibration, external reference targets should be used (Ahokas et al. 2006).

In this study, three MS-ALS datasets acquired at different times of the year are compared. The data used in this study is commercial data, and external reference targets for calibration were not available. In (Matikainen et al. 2017), the object-based mapping of six land cover classes using MS-ALS data and a reference point set was carried out. The achieved overall accuracy was 95.9%. Using similar methods in (Karila et al. 2017), the road detection rate, including also narrow pedestrian and cycle paths, was 84.1%. These studies were based on data acquired in late summer, in August, when trees were in leaf, with the leaves on trees still being green and lower vegetation being green or light brown, depending on the type of vegetation. In this paper, the tests in (Matikainen et al. 2017; Karila et al. 2017) were repeated using two MS-ALS datasets acquired at different times of the growing season: May and June. In early May, trees are leafless and low vegetation is mostly light brown. In June, trees are in leaf and low vegetation is green. In the study area, the leaves turn yellow beginning in September, and, the leaves fall from trees at the latest in October or November. Snow in the winter and fallen leaves in fall covering objects on the ground often disable mapping at other times of the year. The comparison of the multitemporal datasets was based on histogram analyses and object-based classifications trained for each data set separately. In this way, information on the usability of the different data acquisitions for automatic mapping could be obtained.

The objectives of this paper are to study the object-based temporal stability of operational multispectral ALS intensity data for different land cover classes, and to study the effect of seasonal variation on land cover classification and road detection accuracy in a suburban area. Also, the importance of different features in class separation during different

MANUSCRIPT

seasons is reported. This study will provide information to help acquisition planning for automated mapping in the future.

2. Materials

2.1 Study area and reference points

The study area is located in Espoonlahti (60°9′18″N, 24°38′24″E), in southern Finland. It is a suburban area and constantly changing due to urban development. The area includes residential areas, industrial areas, recreational areas and boreal forests. The data used in this study were acquired in 2015 and 2016. In 2015, the thermal growing season started on the 8 April and lasted until the 5 October in the study area. In 2016, it started on the 6 April and lasted until the 24 October (FMI 2018). The start of vegetation period (green-up) in deciduous species occurred between the 1 and 10 May 2016 (SYKE 2018).

A permanent test field of land cover ground control points has been established in the area. These reference points have been used to evaluate the performance of different remotely sensed datasets (Matikainen & Karila 2011; Matikainen et al. 2017). The reference point set was updated to correspond to each data acquisition date. A few points were moved to different locations, and a few others were reclassified or removed. The study area had separate training and test areas, and thus separate training and test point sets. The reference point sets used for land cover classification and the number of points in each set are listed in Table 1. A water mask was derived from the topographic map data, and reference points under the mask have been left out.

[Table 1 near here]

The land cover classes of the reference points used in this study are as follows: building, tree, asphalt, gravel, low vegetation and rocky areas. The tree class includes deciduous and coniferous trees; the most common tree species in the area are Pine (*Pinus*

MANUSCRIPT

silvestris), Spruce (*Picea Abies*) and Birch (*Betula pubescens*). The asphalt class includes roads and parking places with asphalt (and a few with tile) surfaces. The gravel class includes soft, non-vegetated surfaces with different grain sizes (roads, sports fields, beaches). The rocky areas have bare or slightly vegetated surface (typically some moss or patchy grass). The low vegetation class includes grass, meadow, forest floor, vegetable gardens, and low bushes.

For road mapping tests, a different and more extensive set of test points concentrating on different types of roads was used. The road test points (Karila et al. 2017) produced using road database vectors were now updated. Only the overlapping area with the August 2015 land cover classification (Matikainen et al. 2017) results and roads that did not change between data acquisitions were used for collecting the reference points. Finally, a total of 5780 road points remained. The land cover training and test points and road test points are presented in Figure 1.

[Figure 1 near here]

2.2 MS-ALS datasets

The MS-ALS data was acquired using an Optech Titan sensor in cooperation with TerraTec Oy (Helsinki, Finland). The sensor acquires three separate point clouds. The intensity bands of the Optech Titan sensor are infrared 1550 nm (Channel 1, Ch1), near-infrared 1064 nm (Channel 2) and green 532 nm (Channel 3). The channels have different nominal look angles: Ch1: 3.5° forward, Ch2: nadir, Ch3: 7° forward (Ahokas et al. 2016). The acquisition dates were 21 August 2015, 2 May 2016 and 14 June 2016. The main differences between the datasets were the different times of the growing season and especially that in May 2016, trees were not in leaf. The data acquisition parameters and weather conditions are listed in Table 2. Rain did not occur before the data acquisitions.

[Table 2 near here]

MANUSCRIPT

First, a relative radiometric calibration based on range differences was performed on the data (Ahokas et al. 2006; Höfle and Pfeifer 2007; Korpela et al. 2010, Matikainen et al. 2017). Then, overlapping points of different flight lines and some error points were removed using TerraScan (Terrasolid Ltd., Helsinki, Finland) software. Before further analyses, the point clouds were rasterised. Five rasters were generated: the first and only pulse average intensity in a 20 cm grid separately for the three intensity channels (original intensity/100), and a maximum digital surface model (DSM) and minimum DSM from all channels in a 100 cm grid. In addition, a digital terrain model (DTM) produced from the only and last pulse data from August 2015 was used. A more detailed description is available in (Matikainen et al. 2017).

An example of the intensity at three dates is presented in Figure 2. There were still significant differences, after range correction, in the absolute intensity values between the dates (upper row in Figure 2). Possible causes, in addition to seasonal changes in the landscape, are differences in humidity and sensor parameters (Table 2). However, systematic differences in the intensity values between flights were not a problem in our study because the classifier was trained based on the input data separately for each date.

[Figure 2 near here]

To compare the intensity values of the three datasets using histogram analysis (section 4.1), an additional intensity adjustment based on 18 natural and man-made calibration targets (building roofs, parking places, sports fields and beaches) was carried out. The calibration sites were selected so that they included areas with different levels of brightness. Linear fit with scalar adjustment was found between the calibration target mean intensities of a 2016 dataset and the August 2015 dataset. Separate linear models were derived for each of the three intensity channels and for both 2016 datasets. A minimum mean intensity value of the

MANUSCRIPT

reference targets was found for the 2016 datasets, and only the intensity values above the minimum value were adjusted using the linear models.

After the additional intensity adjustment, the intensity levels matched quite well in a visual inspection (Figure 2). However, small intensity differences remained in surfaces expected to remain stable, such as paved roads. Thus, it is impossible to say if they were caused by changes in the conditions or the quality of the selected calibration targets. Since the classification method used in this study took into account the intensity level differences in the training data (i.e., the classifier is trained separately for each dataset), the classification tests presented in this paper (sections 4.2 – 4.5) were carried out on the (not-adjusted) range-corrected data. The adjusted intensity was used for the histogram analyses only.

3. Methods

The three MS-ALS datasets were processed using the method described in (Matikainen et al. 2017). Briefly, it is an object-based approach where each dataset is first processed using the multi-resolution segmentation algorithm (Baatz and Schäpe 2000) in eCognition Software (Trimble Germany GmbH, Munich). Segmentation and feature extraction steps were as follows: (1) First level segmentation was carried out based on MS-ALS MaxDSM raster (segmentation parameters: scale 15, shape 0). (2) The segments were divided into high and low segments based on a mean height threshold of 2.5 m (the standard ceiling height in Finland) from the ground. (3) The low segments were merged and then segmented using the intensity data only (scale 2, shape 0.01 and compactness 0.5). (4) Features based on segment intensity and height were calculated for each high and low segment. The 36 intensity features and 5 DSM features are listed in Table 3.

[Table 3 near here]

In the second stage, all segments and their features were imported to Matlab (The Mathworks, Inc., Natick, MA, USA), where the random forest (RF) (Breiman 2001) method

MANUSCRIPT

was applied to carry out the land cover classification. The high segments were classified using the intensity and DSM features. The low segments were classified using the intensity features only. In Matlab the following steps were carried out. (1) Training segments were selected based on the training points. (2) The 'fitensemble' function with bagging method in Matlab was used for training the RF classifier (i.e. to construct an ensemble of 1000 classification trees). (3) The out-of-bag (OOB) classification error was calculated using the 'oobLoss' function, and the importance of the different features in the classification was estimated using the 'predictorImportance' (Mathworks, 2018) function for the training segments. (4) The ensemble of classification trees generated was used to predict the land cover of all segments. (5) In a simple post-processing step, buildings smaller than 20 m^2 were removed.

Finally, the classification results were validated. The accuracies of the final land cover maps were estimated using the test point sets and geographic information system (GIS) software QGIS (QGIS 2017). The estimation of the road detection rate was based on the three land cover classification results for the gravel and asphalt classes and the road test point set.

To study the distribution of intensity values in the range corrected and intensity adjusted data for all dates, histogram analyses were carried out in Matlab on the training segments. In this case, training points that remained the same from August 2015 to June 2016 were used to define the training segments.

4. Results and Discussion

4.1 Intensity variation of different land cover classes

The histograms (Figures 3 and 4) show the behaviour of the intensity values on different dates in the adjusted and original, non-adjusted (only range-corrected) data. In most cases, the adjusted intensity values matched the intensity values of the first dataset better than the original intensity values. A clear example of this is the histograms of rocky areas in Ch1. The

MANUSCRIPT

benefits of the adjustment, however, are not obvious in all cases, and possible seasonal variations make it difficult to fully evaluate the effectiveness of the adjustment. As expected, for high objects, the difference in intensity stability between natural targets (trees) and man-made objects, such as buildings, was clear. For low objects, there was more variability. It should also be noted that the small number of training objects in the gravel (15) and rocky area (16) classes can make the results from these classes less reliable and stable. Some variation, most likely related to seasonal changes, was also visible in the vegetated classes. For example, the low intensity values in Ch2 were more typical for trees in May than in June and August. To some extent, this also applies to low vegetation.

[Figures 3 and 4 near here]

4.2 OOB errors in land cover classifier training

We carried out a random forest analysis, in which 1000 classification trees were created based on the training data, and, the OOB error rates were estimated based on the training points. OOB errors are listed in Table 4 together with the results from the previous study (Matikainen et al. 2017). In general, the lowest OOB error rates were reported for the August dataset. The classification based on the May dataset had the highest OOB error rates. However, the differences between datasets were quite small.

There were bigger differences in the OOB errors (Table 4) between the dates for low objects than for high objects (building/trees). It can be expected that a Building-Tree classification is simpler to carry out than classification of low objects, which are likely more diverse. June and August basically have similar vegetation cover (leaf-on), however there may be differences in the colour of the vegetation and height of the low vegetation. There were also small differences in the OOB classification error (Table 4) for low objects between June and August. In general, rocky areas are challenging to define because of the presence of

MANUSCRIPT

many low vegetation spots in rocky areas. The small amount of gravel training points and the diversity of the gravel areas affected the results as well.

[Table 4 near here]

4.3 Importance of different features in land cover classification

We estimated the importance of each feature in the land cover classification using the training data. The feature importance values for separating high objects (buildings and trees) and low objects (asphalt, gravel, rocky areas and low vegetation) are presented in Figure 5 for May, June and August. The five most important features for separating the classes are listed in Table 5 for August (Matikainen et al. 2017), May and June. As expected, feature importance (Figure 5) in June was near to that of August (Matikainen et al. 2017), especially for high objects. In general, channel ratios and indices were the most important features. In May, the importance of the features was different for separating the high objects; e.g. the intensity ratios were not as important as in summer and some of the DSM features appeared among the most useful features, unlike in summer. This is likely due to the smaller amount of green vegetation. In all datasets, the texture feature GLCM homogeneity for Ch 2 was important for separating the high objects.

[Figure 5 near here]

[Table 5 near here]

4.4 Land cover classification results

We used the test points to estimate the accuracy of the RF classifier. The classification results for the whole study area and three close-ups are presented in Figure 6 (May) and Figure 7 (June). The corresponding Figure for August was presented in (Matikainen et al. 2017).

The confusion matrices based on the test points are presented in Table 6. The overall accuracy in (Matikainen et al. 2017) of the August dataset was 95.9% and the Kappa 0.95.

MANUSCRIPT

The overall accuracy was slightly higher in June (96.4%, kappa 0.95) and a little lower in May (93.8%, kappa 0.92).

[Table 6 near here]

[Figures 6 and 7 near here]

In a visual inspection of the leaf-off data classification results, we detected misclassifications for open grass areas or meadow classified as gravel. The leaf-off results also contained fewer trees than the summer datasets, and they were replaced by the low object classes. In the May data, there was also seemingly more confusion in the high objects classification than in the summer data. These were likely caused by the lack of green vegetation.

The differences detected in the visual inspection are not supported by the confusion matrices. The testing points were located in the middle of homogenous land cover areas, and many of the visually detected misclassifications were located in the borders of the land cover objects, and thus they were not included in the results. However, the visually detected artefacts were rather small in area and should not significantly affect the quality of the results.

For the May results, the lowest completeness (Table 6) was for gravel, and the second lowest for rocky areas. For the June results, the classification accuracy of gravel was higher, but still the lowest of all the classes. This may have been caused by the variation in gravel surfaces (changes in particle locations, moisture changes) (Kaasalainen et al. 2010), making it easy to confuse them with other classes, especially asphalt. Because of the diversity in the gravel area, more training points are preferred for gravel areas in the future.

The classification results (Figures 6 and 7) can also be used to detect changes. Logging and new-made objects are visible in the results. Change detection based on the multitemporal data and classification results is further analysed in another study (Matikainen et al. 2018).

MANUSCRIPT

4.5 Road detection

We analysed the road detection results using the separate road test point set of 5780 points. The results for August, May and June are presented in Table 7. The road detection rate was highest in May (86.7%) and lowest in June (81.5%). In June, there were difficulties in detecting gravel roads (53.9% vs. 67.6% in May). In a visual inspection, some gravel roads were classified as low vegetation. Many gravel roads in the study area are narrow cycle paths, which may be occluded by trees, and thus the detection rate was lower than the detection rate for asphalt roads. The road detection rates of different road classes are presented in Table 8. The largest differences between dates were found for the narrow roads (cycling paths / driveways). The main roads were detected with a high degree of accuracy in all datasets. These results indicate that leaf-off data (May) is preferred for more complete roads in road detection. However, based on a visual inspection, OOB errors (Table 4) and the confusion matrices (Table 6) asphalt and gravel classes are slightly more confused in May. Therefore, early May does not seem an optimal time for road surface classification.

[Tables 7 and 8 near here]

One question related to the road classification has to do with the small number of gravel training points. This may have had some effect on the road detection results. However, based on our previous study (Karila et al. 2017) with a larger number of road training points, we know that it does not necessarily increase the accuracy significantly. In (Karila et al. 2017), a 2-stage classification for August dataset was carried out: first, a *road/non-road* classification for road detection and then an *asphalt/gravel* classification of the road surface. An expanded set of training points was used (more asphalt and gravel points and more rocky areas in the *non-road* class). Nevertheless, the accuracy was lower than in this study, especially for big roads. However, the results cannot be directly compared to the present study due to different classification strategies.

MANUSCRIPT

4.6 Other seasons and the applicability of the method to other study areas

The typical time for data acquisition flights is during spring or summer. We expect a decrease in the classification accuracy in other seasons in the study area. More colourful vegetation in autumn is likely to cause confusion between the classes used in this study. However, in studies on species classification it may be useful. Snow cover in winter time changes the landscape and bases for mapping completely and makes it impossible to distinguish many land cover classes from each other. For future studies, it would be an interesting research topic to determine whether automated land cover mapping is feasible in other snowless times of the year and estimate the accuracy decrease for autumn datasets or datasets acquired outside the growing season.

The method can be applied in other areas as well. In the parameter selection, local building height should be considered when choosing the threshold for separating high and low objects. The segmentation parameters provided here may be applied as a starting point for MS-ALS raster datasets with similar characteristics, including pixel size and similar ground resolution or point density of the original data. However, depending on the land cover characteristics and the land cover classes, the parameters may need to be adjusted.

5. Conclusions

This paper provides the first results on multitemporal MS-ALS data for land cover classification. All multispectral airborne laser scanning datasets used in this study were suitable for automated suburban land cover classification, regardless of the acquisition date. The automated method was able to find a set of optimum features that separate the selected land use classes (building, tree, low vegetation, asphalt, gravel, rocky area) for each date. The feature importance results can also help in finding good features for class separation when less-automated image interpretation methods are used. Based on this study, the optimal time

MANUSCRIPT

for MS-ALS acquisition for automated suburban land cover mapping under the studied conditions in a hemiboreal zone was summer (June). For automatic road detection, leaf-off conditions were preferred. However, the differences were small and data acquisition for automated mapping with MS-ALS can be carried out both in spring and the summer season.

As already stated in earlier studies considering single-channel ALS intensity, the MS-ALS intensity is also not stable between different acquisitions. In this study, we tested a simple additional intensity adjustment based on calibration sites selected from the data. The benefit of the approach is that external calibration targets are not needed. We achieved a good visual match between the datasets; however, some differences persisted in surfaces expected to remain unchanged. In the future, further studies on calibration and its effect on the classification accuracy of multitemporal datasets are needed.

Funding

This work was supported by the Academy of Finland, under Grants no. 295047 ('Integration of large multisource point cloud and image datasets for adaptive map updating') and no. 307362 (Centre of Excellence in Laser Scanning Research').

References

Ahokas, E., S. Kaasalainen, J. Hyyppä, and J. Suomalainen. 2006. "Calibration of the Optech ALTM 3100 laser scanner intensity data using brightness targets." *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 34: 3-6.

Ahokas, E., J. Hyyppä, X. Yu, X. Liang, L. Matikainen, K. Karila, P. Litkey, A. Kukko, A. Jaakkola, H. Kaartinen, M. Holopainen, and M. Vastaranta. 2016. "Towards automatic single-sensor mapping by multispectral airborne laser scanning." *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B3: 155–162.

Baatz, M., and A. Schäpe. 2000. "Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation." In: Strobl, J., Blaschke, T.,

MANUSCRIPT

Griesebner, G. (Eds.), *Angewandte Geographische Informationsverarbeitung XII: Beiträge Zum AGIT-Symposium Salzburg 2000.* Wichmann, Heidelberg, pp. 12–23

Bakuła, K., P. Kupidura, and L. Jełowicki. 2016. "Testing of land cover classification from multispectral airborne laser scanning data," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B7: 161-169.

Breiman, L. 2001. "Random forests." *Machine Learning* 45 (1): 5–32.

Briese, C., M. Pfennigbauer, A. Ullrich, and M. Doneus. 2013. "Multi-wavelength airborne laser scanning for archaeological prospection," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-5/W2: 119-124.

Budei, B.C., B. St-Onge, C. Hopkinson, and F.-A. Audet. 2018. "Identifying the genus or species of individual trees using a three-wavelength airborne lidar system," *Remote Sensing of Environment* 204: 632-647.

Clark, M.L. 2017. "Comparison of simulated hyperspectral HyspIRI and multispectral Landsat 8 and Sentinel-2 imagery for multi-seasonal, regional land-cover mapping," *Remote Sensing of Environment* 200: 311-325.

Fernandez-Diaz, J.C., W.E. Carter, C. Glennie, R.L. Shrestha, Z. Pan, N. Ekhtari, A. Singhania, D. Hauser, and M. Sartori. 2016. "Capability assessment and performance metrics for the Titan multispectral mapping lidar." *Remote Sensing* 8 (11): 936.

FMI (Finnish Meteorological Institute) 2018. http://ilmatieteenlaitos.fi/terminen-kasvukausi (in Finnish). (accessed on 31 May 2018)

Gamba, P., and B. Houshmand. 2002. "Joint analysis of SAR, LIDAR and aerial imagery for simultaneous extraction of land cover, DTM and 3D shape of buildings," *International Journal of Remote Sensing* 23 (20): 4439-4450.

Guo, L., N. Chehata, C. Mallet, S. Boukir. 2011. "Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests," *ISPRS Journal of Photogrammetry and Remote Sensing*, 66: 56–66.

Heinl, M., and U. Tappeiner. 2012. "The benefits of considering land cover seasonality in multi-spectral image classification," *Journal of Land Use Science* 7(1): 1-19.

Höfle, B., and N. Pfeifer. 2007. "Correction of laser scanning intensity data: Data and model-driven approaches," *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (6): 415–433.

MANUSCRIPT

Hopkinson, C., L. Chasmer, C. Gynan, C. Mahoney, and M. Sitar. 2016. "Multisensor and multispectral LiDAR characterization and classification of a forest environment," *Canadian Journal of Remote Sensing* 42 (5): 501-520.

Huang, M.-J., S.-W. Shyue, L.- H. Lee, and C.-C. Kao. 2008. "A knowledge-based approach to urban feature classification using aerial imagery with lidar data." *Photogrammetric Engineering and Remote Sensing* 74: 1473-1485.

Hug, C., A. Wehr. 1997. Detecting and identifying topographic objects in laser altimeter data *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 32 (Part 3–4/W2): 19-26

QGIS Development Team 2017. QGIS Geographic Information System. Open Source Geospatial Foundation. http://qgis.osgeo.org (accessed 30 May, 2017).

Kaasalainen, S., H. Niittymäki, A. Krooks, K. Koch, H. Kaartinen, A. Vain, and H. Hyyppä. 2010. "Effect of Target Moisture on Laser Scanner Intensity." *IEEE Transactions on Geoscience and Remote Sensing* 48 (4): 2128-2136.

Kaasalainen, S., U. Pyysalo, A. Krooks, A. Vain, A. Kukko, J. Hyyppä, and M. Kaasalainen. 2011. "Absolute Radiometric Calibration of ALS Intensity Data: Effects on Accuracy and Target Classification." *Sensors* 11: 10586-10602.

Karila, K., L. Matikainen, E. Puttonen, and J. Hyyppä. 2017. "Feasibility of multispectral airborne laser scanning data for road mapping." *IEEE Geoscience and Remote Sensing Letters* 14 (3): 294-298.

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.

Korpela, I., H.O. Ørka, V. Heikkinen, T. Tokola, and J. Hyyppä. 2010. "Range- and AGC normalization of LIDAR intensity data for vegetation classification," *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (4): 369-379.

Leigh, H.W., and L.A. Magruder. 2016. "Using dual-wavelength, full-waveform airborne lidar for surface classification and vegetation characterization," *Journal of Applied Remote Sensing* 10 (4): 045001.

Mathworks, 2018. Online Documentation for Statistics and Machine Learning Toolbox, Version R2017b.

MANUSCRIPT

Matikainen, L., and K. Karila. 2011. “Segment-Based Land Cover Mapping of a Suburban Area—Comparison of High-Resolution Remotely Sensed Datasets Using Classification Trees and Test Field Points.” *Remote Sensing* 3: 1777-1804.

Matikainen, L., K. Karila, J. Hyyppä, P. Litkey, E. Puttonen, and E. Ahokas. 2017.” Object-based analysis of multispectral airborne laser scanner data for land cover classification and map updating.” *ISPRS Journal of Photogrammetry and Remote Sensing* 128: 298-313.

Matikainen, L., M. Pandžić, F. Li, K. Karila, J. Hyyppä, P. Litkey, A. Kukko, M. Lehtomäki, M. Karjalainen, and E. Puttonen. 2018. “Towards integration of multitemporal multispectral airborne laser scanning, Sentinel-2 satellite images and mobile laser scanning in updating of topographic databases.” Submitted manuscript.

Morsy, S., A. Shaker, and A. El-Rabbany. 2017. “Multispectral LiDAR data for land cover classification of urban areas.” *Sensors* 17: 958.

Nabucet, J., L. Hubert-Moy, T. Corpetti, P. Launeau, D. Lague, C. Michon, and H. Quenol. 2016. “Evaluation of bispectral LiDAR data for urban vegetation mapping.” In *Proceedings of SPIE 10008*. doi: 10.1117/12.2241731.

Salah, M., J. Trinder, and A. Shaker. 2009. “Evaluation of self-organizing map classifier for building detection from lidar data and multispectral aerial images.” *Journal of Spatial Science* 54:15–34.

Sugumaran, R., M.K. Pavuluri, and D. Zerr. 2003. “The use of high-resolution imagery for identification of urban climax forest species using traditional and rule-based classification approach.” *IEEE Transactions on Geoscience and Remote Sensing* 41: 1933-1939.

SYKE (Finnish Environment Institute) 2018. “Phenology: Start of vegetation period in deciduous species 2001-2016.” http://www.syke.fi/en-US/Open_information/Spatial_datasets (accessed on 31 May 2018)

Teo, T.-A., and H.-M. Wu. 2017. “Analysis of land cover classification using multi-wavelength LiDAR system.” *Applied Sciences* 7: 663.

Wichmann, V., M. Bremer, J. Lindenberger, M. Rutzinger, C. Georges, and F. Petrini-Monteferri. 2015. “Evaluating the potential of multispectral airborne lidar for topographic mapping and land cover classification.” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W5: 113-119.

MANUSCRIPT

Yan, W.Y., and A. Shaker. 2014. "Radiometric Correction and Normalization of Airborne LiDAR Intensity Data for Improving Land-Cover Classification." *IEEE Transactions on Geoscience and Remote Sensing* 52: 7658-7673.

Voss, M., and R. Sugumaran. 2008. "Seasonal Effect on Tree Species Classification in an Urban Environment Using Hyperspectral Data, LiDAR, and an Object- Oriented Approach." *Sensors* 8: 3020-3036.

Yan, W.Y., A. Shaker, and N. El-Ashmawy, 2015. "Urban land cover classification using airborne LiDAR data: A review," *Remote Sensing of Environment* 158: 295-310.

Yan, W. Y. and A. Shaker. 2017. "Correction of overlapping multispectral lidar intensity data: polynomial approximation of range and angle effects." *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-3/W1: 177-182. doi.org/10.5194/isprs-archives-XLII-3-W1-177-2017.

Yu, X., J. Hyyppä, P. Litkey, H. Kaartinen, M. Vastaranta, and M. Holopainen. 2017. "Single-sensor solution to tree species classification using multispectral airborne laser scanning," *Remote Sensing* 9(2): 108.

Zhou, W. 2013. "An Object-Based Approach for Urban Land Cover Classification: Integrating LiDAR Height and Intensity Data," *IEEE Geoscience and Remote Sensing Letters* 10 (4): 928-931.

MANUSCRIPT

Table 1. The number of land cover reference points.

	August 2015		May 2016		June 2016	
Class	**Training points**	**Test points**	**Training points**	**Test points**	**Training points**	**Test points**
Building	88	130	88	130	88	130
Tree	83	142	83	141	85	141
Asphalt	62	157	63	157	63	157
Gravel	15	15	15	15	15	15
Rocky area	16	44	16	44	16	44
Low vegetation	68	76	67	76	65	76
Total	**332**	**566**	**332**	**565**	**332**	**565**

MANUSCRIPT

Table 2. The Optech Titan MS-ALS datasets.

	Date & time (UTC)	**Flying height (m)**	**Laser pulse rate (kHz)**	**Point density Ch1, Ch2, Ch3 (points per m^2)**	**Humidity (%)**	**Temp. (°C)**	**Trees in leaf**
August	21.8.2015 17:01-17:57	650	200	9, 9, 8	93	14	Yes
May	2.5.2016 8:12-9:12	700	300	12, 14, 10	34	16	No
June	14.6.2016 6:26-7:15	700	300	11, 13, 11	48	16	Yes

MANUSCRIPT

Table 3. Intensity and DSM (height)* features used in the study.

Feature type	**Data and additional details**
Brightness	the mean value of the mean intensity values in different channels
Intensity mean, 10th percentile of intensity (Q10), Q25, Q50, Q75, Q90,	ch1, ch2, ch3
Ratio to all	ch1, ch2, ch3 (the mean intensity in one channel divided by the sum of the mean intensity values in all channels)
Standard deviation	ch1, ch2, ch3, minDSM*, maxDSM*
Grey-level co-occurrence matrix (GLCM) homogeneity	ch1, ch2, ch3, minDSM*, maxDSM*
Ratios of two channels	ch1/ch3, ch1/ch2, ch2/ch3
Indices	pseudo NDVI (normalized difference vegetation index): (Mean Ch2 – Mean Ch 3)/(Mean Ch2 + Mean Ch3) pseudo NDBI (normalized difference built-up index): (Mean Ch1 – Mean Ch2)/(Mean Ch1 + Mean Ch2)
Differences	Q90-Q10 (Ch1, Ch2, Ch3), maxDSM-minDSM*

MANUSCRIPT

Table 4. The OOB errors of the random forest land cover classifier. August 2015 results originally presented by Matikainen et al. (2017).

	High objects (intensity features)	**High objects (DSM features)**	**High objects (DSM + intensity features)**	**Low objects (intensity features only)**	**Road-like surfaces/rocky areas/low vegetation (intensity features only)**	**Asphalt/gravel (intensity features only)**
August 2015	0.00592	0.0178	0	0.0314	0.0189	0.0263
May 2016	0.012	0.012	0	0.0881	0.0818	0.0533
June 2016	0.0058	0.0234	0.0058	0.0506	0.0506	0.0267

MANUSCRIPT

Table 5. The five most important features for the May, June and August datasets in high object classification (top) and low object classification (bottom). August 2015 results originally presented by Matikainen et al. (2017).

August 2015: high objects	May 2016: high objects	June 2016: high objects
Ratio Ch2	MaxDSM - MinDSM	Ratio Ch2
GLCM hom. Ch2	GLCM hom. Ch2	GLCM hom. Ch2
PseudoNDVI	Q10 Ch3	Ch2/Ch3
Ch2/Ch3	Std MaxDSM	Ratio Ch3
Ch1/Ch2	Q25 Ch3	PseudoNDVI

August 2015: low objects	May 2016: low objects	June 2016: low objects
Ch1/Ch2	Ch1/Ch3	Ratio Ch3
PseudoNDBI	Ratio Ch3	Ch1/Ch3
Ratio Ch3	Ch2/Ch3	PseudoNDVI
Ch1/Ch3	PseudoNDVI	Ch2/Ch3
Ratio Ch2	Ratio Ch2	Ratio Ch2

MANUSCRIPT

Table 6. The confusion matrices for land cover classification for May 2016, June 2016 and August 2015. August 2015 results originally presented by Matikainen et al. (2017)

August 2015		Reference data							
		Building	Tree	Asphalt	Gravel	Rocky area	Low veg.	Total	Correctness
Classification result	Building	**125**	0	0	0	0	0	125	100.00 %
	Tree	3	**142**	0	0	0	0	145	97.90 %
	Asphalt	1	0	**149**	3	0	0	153	97.40 %
	Gravel	1	0	5	**12**	0	0	18	66.70 %
	Rocky area	0	0	1	0	**40**	3	44	90.90 %
	Low veg.	0	1	2	0	4	**73**	79	92.40 %
	Total	130	142	157	15	44	76	564	
	Completeness	96.20 %	100 %	94.90 %	80.00 %	90.90 %	96.10 %		
		Kappa	0.95				Overall accuracy		95.90 %

May 2016		Reference data							
		Building	Tree	Asphalt	Gravel	Rocky area	Low veg.	Total	Correctness
Classification result	Building	**127**	0	1	1	0	0	129	98.40 %
	Tree	1	**136**	0	0	0	0	137	99.30 %
	Asphalt	2	0	**148**	5	0	0	155	95.50 %
	Gravel	0	0	7	**9**	1	1	18	50.00 %
	Rocky area	0	0	0	0	**34**	1	35	97.10 %
	Low veg.	0	5	1	0	9	**74**	89	83.10 %
	Total	130	141	157	15	44	76	563	
	Completeness	97.70 %	96.50 %	94.30 %	60.00 %	77.30 %	97.40 %		
		Kappa	0.92				Overall accuracy		93.80 %

June 2016		Reference data							
		Building	Tree	Asphalt	Gravel	Rocky area	Low veg.	Total	Correctness
Classification result	Building	**127**	0	0	0	0	0	127	100 %
	Tree	1	**140**	0	0	0	0	141	99.30 %
	Asphalt	2	0	**154**	3	0	0	159	96.90 %
	Gravel	0	0	2	**12**	1	1	16	75.00 %
	Rocky area	0	0	0	0	**35**	0	35	100 %
	Low veg.	0	1	1	0	8	**75**	85	88.20 %
	Total	130	141	157	15	44	76	563	
	Completeness	97.70 %	99.30 %	98.10 %	80.00 %	79.50 %	98.70 %		
		Kappa	0.95				Overall accuracy		96.40 %

MANUSCRIPT

Table 7. The road detection rates based on the road test points for all datasets

	August 2015	**May 2016**	**June 2016**	**# test points**
Gravel road detected	61.6 %	67.6 %	53.9 %	1515
Asphalt road detected	92.0 %	93.5 %	91.2 %	4265
Total	**84.1 %**	**86.7 %**	**81.5 %**	**5780**

MANUSCRIPT

Table 8. Road detection rates for road classes

Road class	August 2015	May 2016	June 2016	# test points
Expressway	98.3 %	99.3 %	96.6 %	417
Road, 2 lanes, 5 - 8 m	96.4 %	96.9 %	96.6 %	745
Road, 1 lane, 3 - 5 m	91.9 %	93.5 %	89.8 %	2314
Cycle way / Driveway < 3 m	69.7 %	74.4 %	65.4 %	2304

MANUSCRIPT

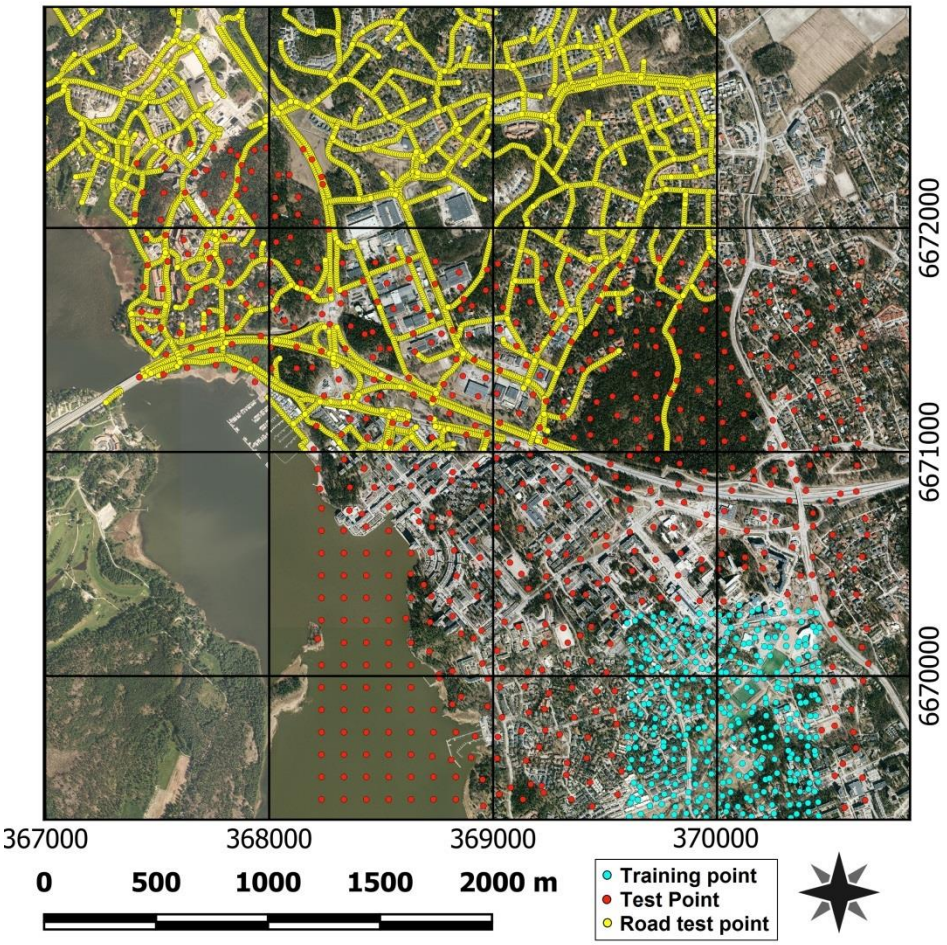

Figure 1. The training (cyan) and test (red) points in the study area. An aerial ortho image is shown in the background. The road test points (yellow) were extracted from road vectors in the National Land Survey of Finland Topographic database 2015. Water areas were excluded from the study. Aerial ortho image ©National Land Survey 2013.

MANUSCRIPT

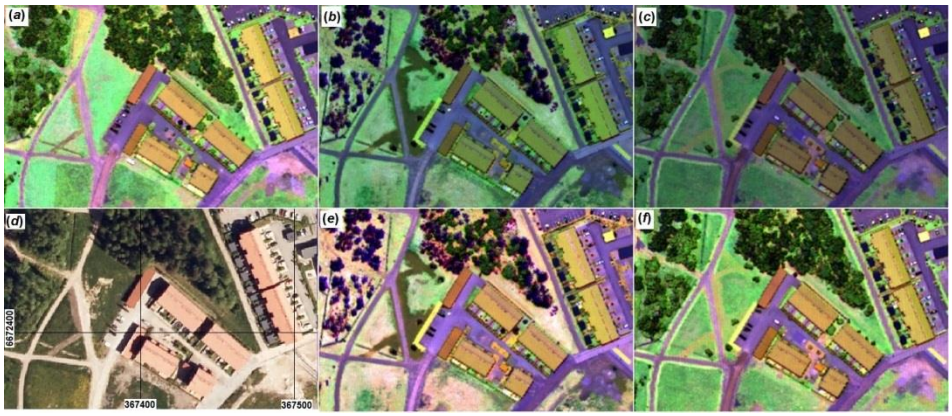

Figure 2. Original range-corrected first and only pulse intensity images for a subarea: (*a*) August 2015, (*b*) May 2016, and (*c*) June 2016). (*d*) Aerial ortho image (the image acquisition date is different from MS-ALS acquisition date, © National Land Survey, 2013). Adjusted intensity images: (*e*) May 2016 and (*f*) June 2016. All colours matched to the August 2015 intensity image (*a*). Red: Ch1; Green: Ch2; Blue: Ch3.

MANUSCRIPT

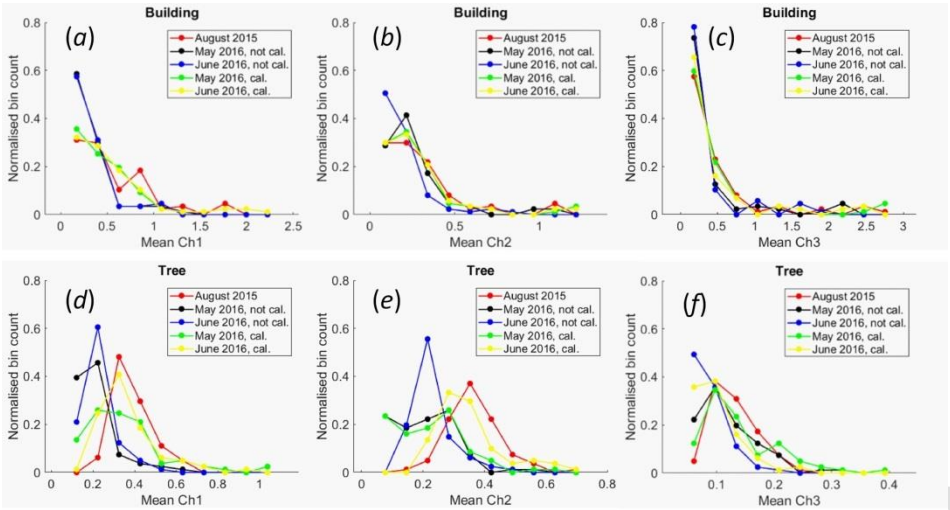

Figure 3. Histograms of *building* (*a*) – (*c*) and *tree* (*d*) – (*f*) training segments in different intensity images. Mean intensity values of the segments were used to calculate the histograms. Not cal. is the range-corrected data and cal. is the range-corrected data with an additional intensity adjustment.

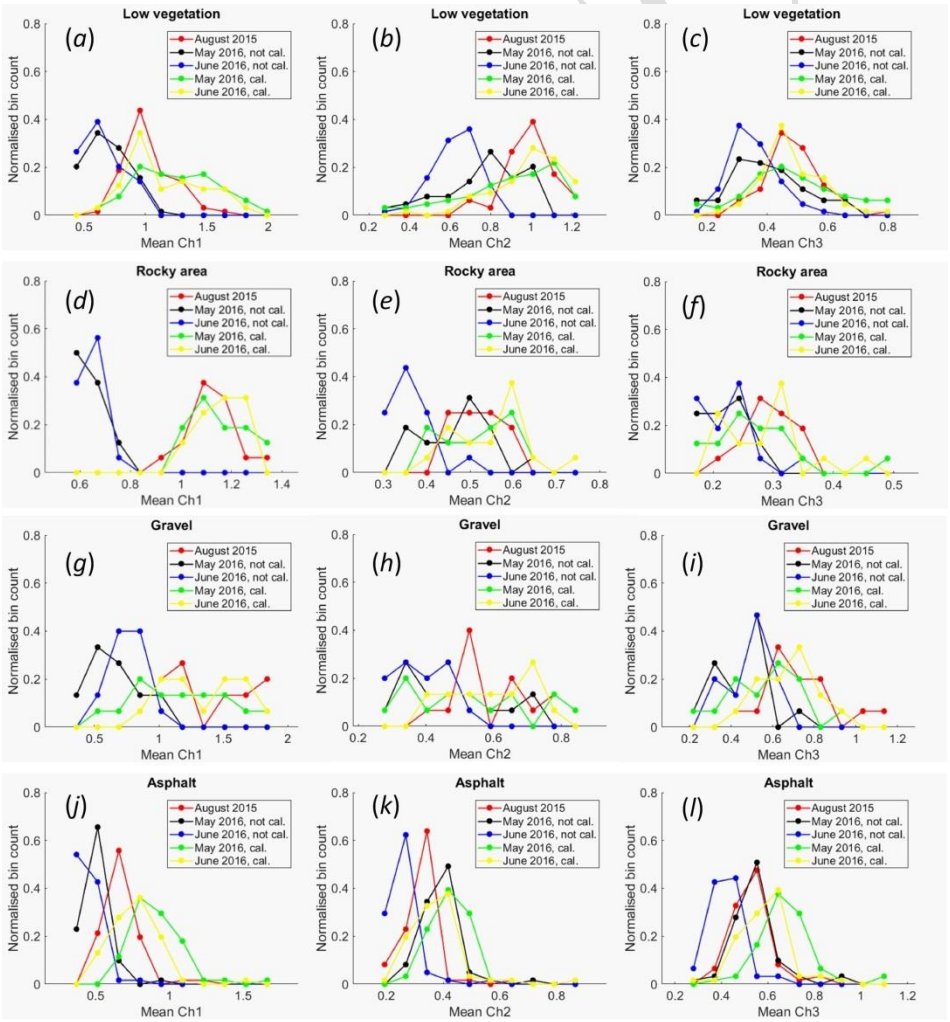

Figure 4. Histograms of *low vegetation* (*a*) – (*c*), *rocky area* (*d*) – (*f*), *gravel* (*g*) – (*i*), and *asphalt* (*j*)-(*l*) training segments in different intensity images. Mean intensity values of the segments were used to calculate the histograms. Not cal. is the range-corrected data and cal. is the range-corrected data with an additional intensity adjustment.

MANUSCRIPT

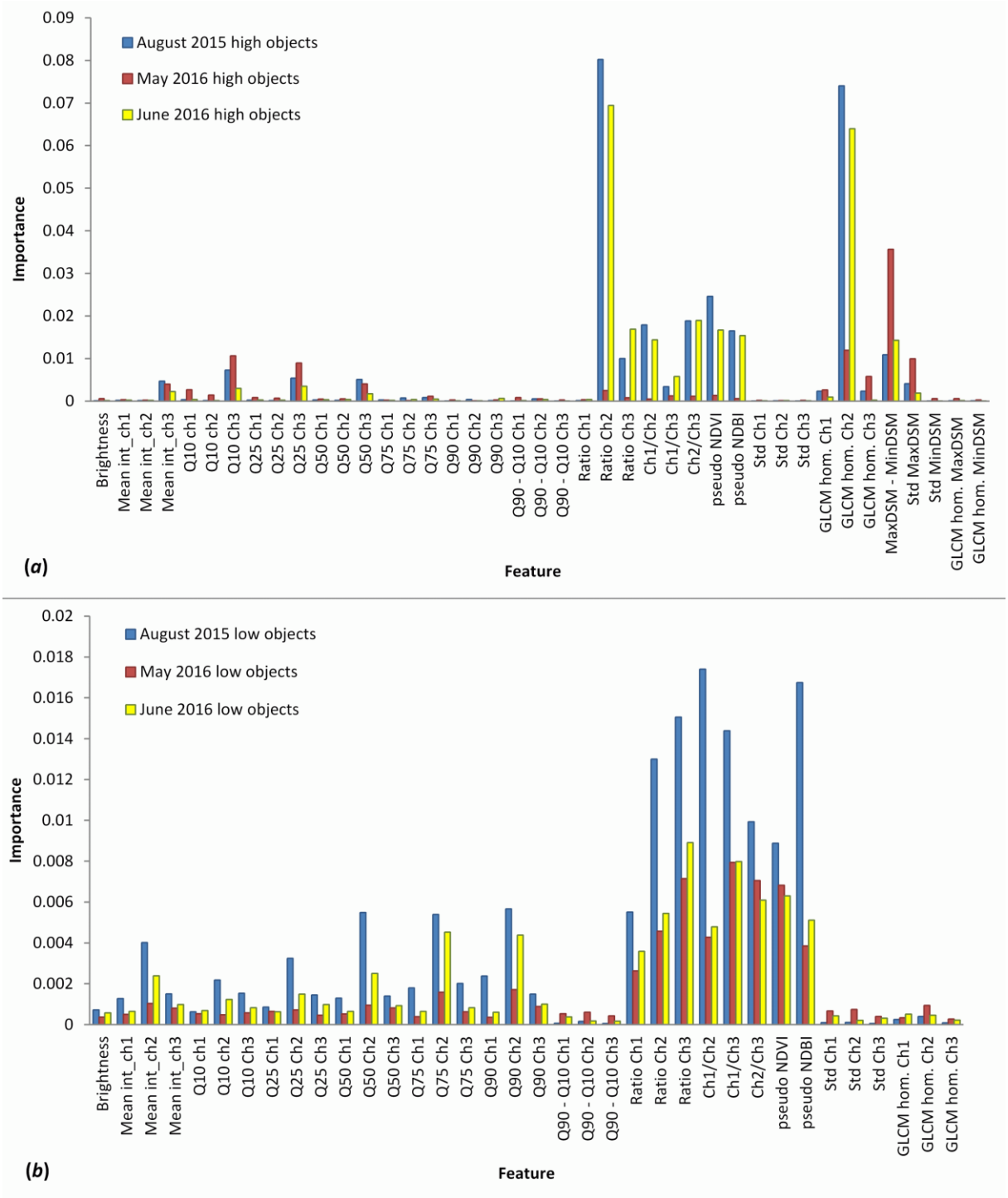

Figure 5. Feature importance for high objects (*a*) and low objects (*b*) for the three datasets. QX = intensity percentile X%. August 2015 results originally presented by Matikainen et al. (2017).

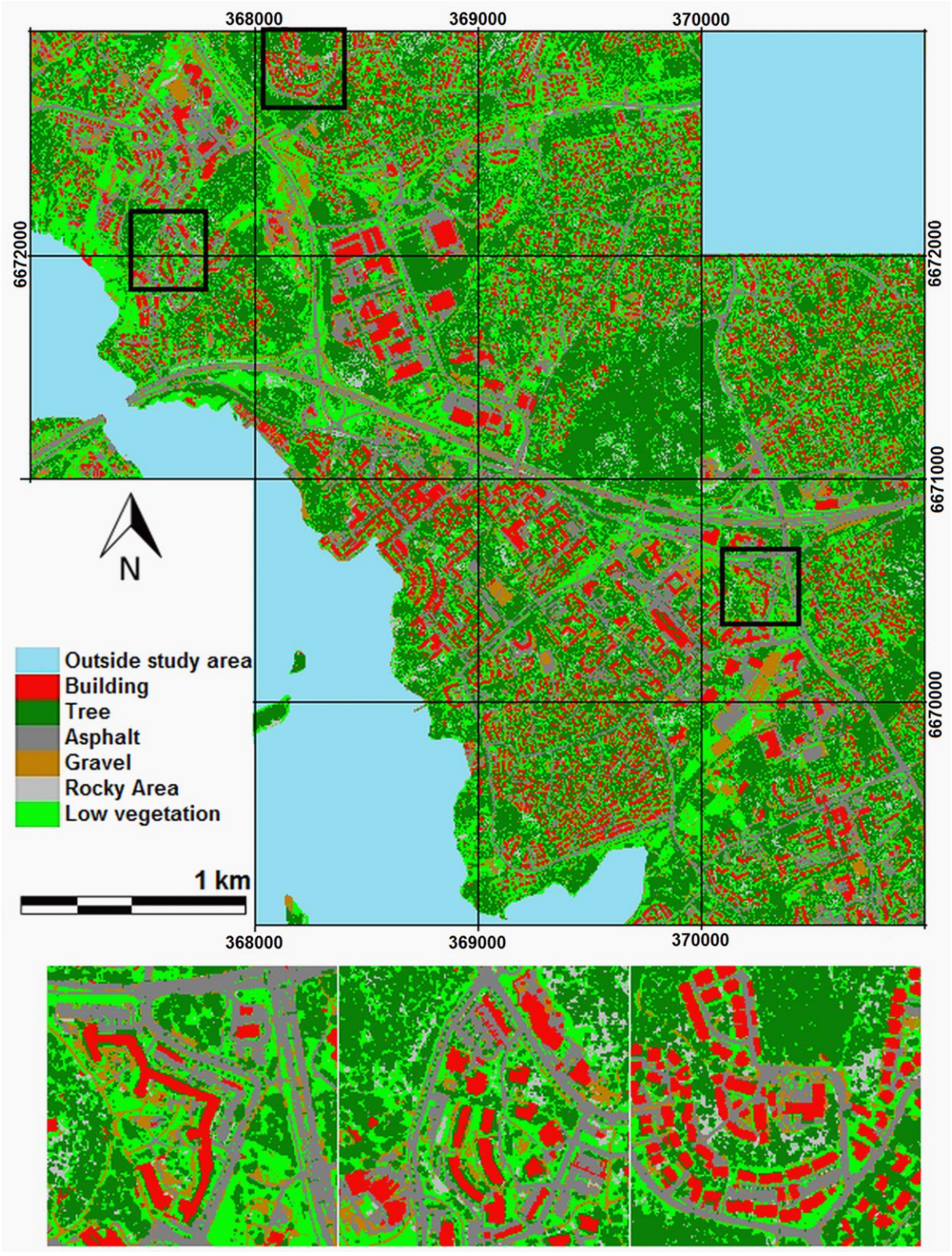

Figure 6. May land cover classification result for the whole study area and three close-ups (black rectangles). The water mask contains data from the National Land Survey of Finland Topographic Database 2015.

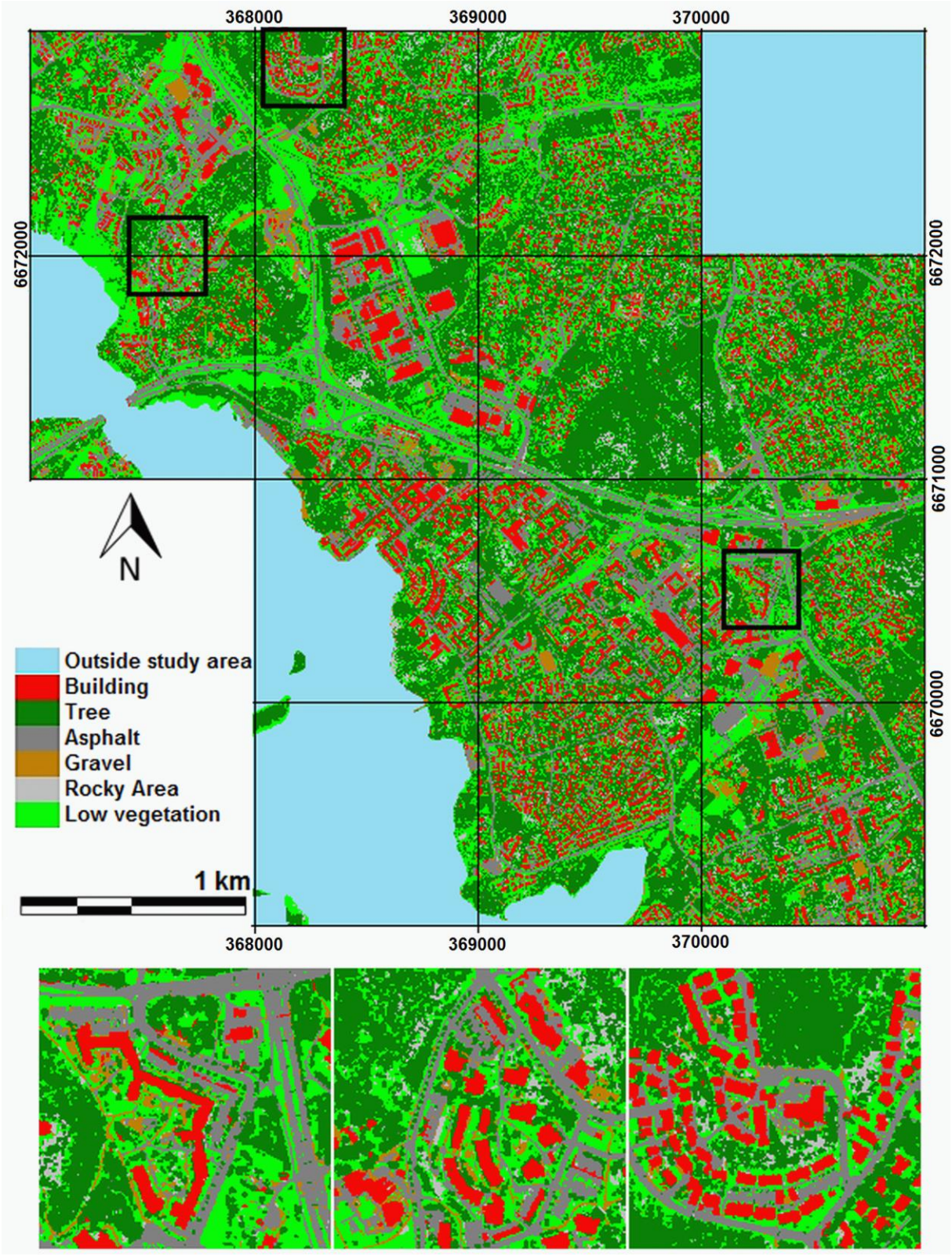

Figure 7. June land cover classification result for the whole study area and three close-ups. The water mask contains data from the National Land Survey of Finland Topographic Database 2015.