

# AIRBORNE MULTISPECTRAL LIDAR DATA FOR LAND-COVER CLASSIFICATION AND LAND/WATER MAPPING USING DIFFERENT SPECTRAL INDEXES

S. Morsy<sup>a,\*</sup>, A. Shaker<sup>a</sup>, A. El-Rabbany<sup>a</sup>, P. E. LaRocque<sup>b</sup>

<sup>a</sup> Department of Civil Engineering, Ryerson University, 350 Victoria St, Toronto, ON M5B 2K3, Canada - (salem.morsy, ahmed.shaker, rabbany)@ryerson.ca

<sup>b</sup> Teledyne Optech, 300 Interchange Way, Concord, ON L4K 5Z8, Canada - Paul.LaRocque@teledyneoptech.com

Commission III, WG III/2

**KEY WORDS:** LiDAR, multispectral, MDVI, NDWI, vegetation, built-up areas, land-water mapping

## ABSTRACT:

Airborne Light Detection And Ranging (LiDAR) data is widely used in remote sensing applications, such as topographic and land-water mapping. Recently, airborne multispectral LiDAR sensors, which acquire data at different wavelengths, are available, thus allows recording a diversity of intensity values from different land features. In this study, three normalized difference feature indexes (NDFI), for vegetation, water, and built-up area mapping, were evaluated. The NDFIs namely,  $NDFI_{G-NIR}$ ,  $NDFI_{G-MIR}$ , and  $NDFI_{NIR-MIR}$  were calculated using data collected at three wavelengths; green: 532 nm, near-infrared (NIR): 1064 nm, and mid-infrared (MIR): 1550 nm by the world's first airborne multispectral LiDAR sensor "Optech Titan". The Jenks natural breaks optimization method was used to determine the threshold values for each NDFI, in order to cluster the 3D point data into two classes (water and land or vegetation and built-up area). Two sites at Scarborough, Ontario, Canada were tested to evaluate the performance of the NDFIs for land-water, vegetation, and built-up area mapping. The use of the three NDFIs succeeded to discriminate vegetation from built-up areas with an overall accuracy of 92.51%. Based on the classification results, it is suggested to use  $NDFI_{G-MIR}$  and  $NDFI_{NIR-MIR}$  for vegetation and built-up areas extraction, respectively. The clustering results show that the direct use of NDFIs for land-water mapping has low performance. Therefore, the clustered classes, based on the NDFIs, are constrained by the recorded number of returns from different wavelengths, thus the overall accuracy is improved to 96.98%.

## 1. INTRODUCTION

The spectral vegetation, water, or built-up index is a single number derived for each pixel from an arithmetic operation on two spectral bands. An appropriate threshold of the index is then established to contrast the reflectance of one feature from other land cover features (e.g., vegetation from built-up area) based on the spectral characteristics. The selection of bands for any spectral index is based on the interaction of features with different wavelengths. Following are examples to existing indexes that have been used to extract different features using multi-spectral bands.

The Normalized Difference Vegetation Index (NDVI) was first proposed by (Rouse et al., 1974) as following:

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \quad (1)$$

where  $\rho_{NIR}$  and  $\rho_{red}$  are the reflectance from the NIR and red bands, respectively.

The basic idea of the NDVI is that the chlorophyll in green plants absorbs red light portion of the spectrum, while NIR light portion is reflected or scattered. As a result, vegetation has low reflectance from red light and high reflectance from NIR light, and hence high NDVI values. The NDVI is used for vegetation extraction (Chen et al., 2009), tree canopy mapping (MacFaden

et al., 2012), and objects classification into vegetation and built-up areas (Thanapura et al., 2007; Zhou et al., 2009; Hartfield et al., 2011).

The Normalized Difference Water Index (NDWI) was first defined by McFeeters (1996) as following:

$$NDWI = \frac{\rho_{green} - \rho_{NIR}}{\rho_{green} + \rho_{NIR}} \quad (2)$$

where  $\rho_{green}$  is the reflectance from the green band.

McFeeters (1996) has selected the green wavelength to maximize the reflectance from water bodies, and the NIR wavelength to minimize the low reflectance from water bodies as well as take the advantage of high reflectance from vegetation and soil features. Based on that, the NDWI has been calculated to discriminate water bodies from vegetation and soil features. Xu (2006) has used the shortwave-infrared (SWIR) instead of NIR to enhance McFeeters's NDWI in distinguishing water bodies from built-up features as well as vegetation and soil features. Xu's MNDWI is defined as following:

$$MNDWI = \frac{\rho_{green} - \rho_{SWIR}}{\rho_{green} + \rho_{SWIR}} \quad (3)$$

where  $\rho_{SWIR}$  is the reflectance from the SWIR band.

\* Corresponding author

Gao (1996) has suggested a new NDWI that should be considered as an independent vegetation index for estimating water content of vegetation canopy. Although vegetation canopies have high reflectance from NIR and SWIR wavelengths between 900–2500 nm, liquid water absorption in SWIR region (1500–2500 nm) is significantly stronger than that in NIR region (900–1300 nm). Gao’s NDWI is defined as:

$$NDWI_{Gao} = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}} \quad (4)$$

Also, the aforementioned equation (Eq.4) has been defined as normalized difference built-up index (NDBI) (Zha et al., 2003; Xu, 2007) and normalized difference soil index (NDSI) (Rogers and Kearney, 2004) to extract built-up areas and soil features, respectively.

Over the past years, remotely sensed satellite data have been employed to extract boundaries of water bodies, vegetation, and built-up areas. Few studies attempted to use NIR band from LiDAR data with red band from aerial images to derive NDVI for vegetation classification (trees and/or grass) in urban areas (Huang et al., 2008; Guan et al., 2013). Recently, the NDVI is calculated for vegetation applications using multispectral LiDAR data obtained from either laboratory-based multispectral LiDAR systems or terrestrial laser scanning (TLS). Woodhouse et al. (2011) used laboratory-based measurements from a tunable laser, operates at four wavelengths (531, 550, 660, and 780 nm). The NDVI and photochemical reflectance index (PRI) were then used to measure plant physiology. Shuo et al. (2015) used a laboratory-based prototype of multispectral LiDAR, designed by (Wei et al., 2012), to record intensity values at four wavelengths (556, 670, 700, and 780 nm). The physiology of the canopy was then studied using three vegetation indexes, and achieved classification accuracy above 90%. Puttonen et al. (2015) used a TLS from the Finnish Geodetic Institute (FGI), called Hyperspectral LiDAR system (HSL). The system was tested in an outdoor experiment using seven wavelength bands ranging from 500 to 980 nm to detect man-made targets from vegetation based on their spectral response. Four vegetation indexes were used in classification and the results achieved an overall accuracy up to 91.0%.

In 2014, Teledyne Optech has launched the world’s first airborne multispectral LiDAR sensor “Optech Titan”. The sensor operates simultaneously at three wavelengths and collects data in three channels: NIR (1064nm) in C2 at 0° nadir looking, MIR (1550nm) in C1 at 3.5° forward looking, and green (532nm) in C3 at 7° forward looking. Specifications of Optech Titan sensor are provided in Table 1. Wichmann et al., (2015) analysed a dataset, acquired by Optech Titan, covering the city of Oshawa. Conventional geometrical classification and mapping procedures were used to classify the terrain into four classes: ground, buildings, mid vegetation, and high vegetation. Spectral patterns for different classes were analysed and showed that the intensity values could be potentially used in classification.

Parameter	Specification
<b>Wavelength</b>	Channel 1 = 1550 nm, Channel 2 = 1064 nm, Channel 3 = 532 nm
<b>Altitude</b>	Topographic: 300-2000 m, all channels Bathymetric: 300-600 m, channel 3

<b>Scan Angle</b>	Programmable; 0 - 60° max
<b>Beam Divergence</b>	Channel 1 & 2 = 0.35 mrad , Channel 3 = 0.7 mrad
<b>Pulse Repetition Frequency (PRF)</b>	50 - 300 kHz/channel; 900 kHz total
<b>Scan Frequency</b>	Programmable; 0 - 210 Hz
<b>Swath width</b>	0 - 115% of AGL
<b>Point Density</b>	Bathymetric: >15 pts/m <sup>2</sup> Topographic: >45 pts/m <sup>2</sup>

Table 1. Optech Titan sensor specifications

This paper aims to evaluate the use of different spectral indexes, as an automatic unsupervised classification method, derived from airborne multispectral LiDAR data, for land-water, vegetation, and built-up area mapping. In this study, three different indexes: 1) the NDFI<sub>G-NIR</sub>, 2) NDFI<sub>G-MIR</sub>, and 3) NDFI<sub>NIR-MIR</sub> are tested based on the three different wavelengths (green (532 nm), NIR (1064 nm), and MIR (1550 nm)) produced by Optech Titan. An automatic 3D point clustering based on spectral indexes is presented, and the recorded returns from different wavelengths are used with the NDFIs for land-water mapping.

## 2. STUDY AREAS AND DATASETS

An urban study area, located near Lake Ontario in Scarborough, Ontario, Canada, was used in this paper. The study area covers variety of land features, such as built-up areas (buildings, roads, and parking lots), high vegetation (trees) and low vegetation (shrubs and grass), as well as part of Lake Ontario. A flight mission, to acquire LiDAR data for three strips (L1, L2, and L3), was conducted on September 3<sup>rd</sup>, 2014. Two subsets of LiDAR data were clipped from L1 (Site I), Figure 1, and L3 (Site II), Figure 2, with the dimension of 600m x 200m each, for the experimental testing. Site I includes land features and the water body so that the NDFIs could be evaluated for land/water mapping, while Site II includes different land features to evaluate the NDFIs for vegetation and built-up area discrimination.

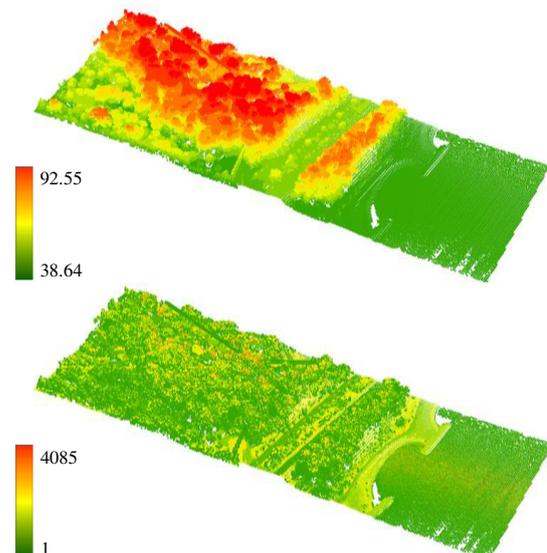


Figure 1. LiDAR point clouds from C2 for Site I (upper: height and lower: intensity)

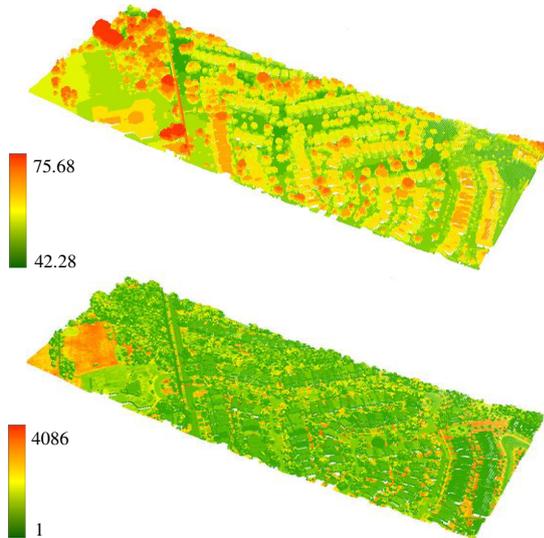


Figure 2. LiDAR point clouds from C2 for Site II (upper: height and lower: intensity)

Optech Titan sensor acquired LiDAR data for L1 and L3 in the three channels at ~420m altitude, ±15° scan angle, 200 kHz/channel PRF and 40 Hz scan frequency. The tested LiDAR data of Site I contains about 1755530, 2013832, and 2256815 points from C1, C2, and C3, respectively, while Site II contains about 1874916, 2022127, and 2036271 points from C1, C2, and C3, respectively. The difference in number of points depends on the interaction of each feature with different wavelengths (e.g., reflection from water surface and/or seabed, and greenness of the vegetation). A Google® image was used to validate the classification results using number of reference points, obtained from selected polygons for different land covers, Figure 3.



Figure 3. Google image showing Site I and Site II

Since the number of points is not the same from each channel, the maximum number of points of the two input channels per each index was used in processing and displaying, in order to obtain highest point density. The points from green channel were used for  $NDFI_{G-NIR}$  and  $NDFI_{G-MIR}$ , while the points from NIR channel were used for  $NDFI_{NIR-MIR}$ . The number of reference points for results validation, for each feature and total number, per each index are provided in Table 2 for the two sites.

Site I			
Index	Land	Water	Total
$NDFI_{G-NIR} / NDFI_{G-MIR}$	90151	63885	154036
$NDFI_{NIR-MIR}$	99018	28792	127810
Site II			
Index	Vegetation	Built-up	Total
$NDFI_{G-NIR} / NDFI_{G-MIR}$	58769	29999	88768
$NDFI_{NIR-MIR}$	64544	29950	94494

Table 2. Reference points for Site I and Site II

### 3. METHODOLOGY

The overall workflow of this study is shown in Figure 4. The intensity values from the green, NIR, and MIR wavelengths of multispectral LiDAR sensor were employed to form  $NDFI_{G-NIR}$ ,  $NDFI_{G-MIR}$ , and  $NDFI_{NIR-MIR}$  as following:

$$NDFI_{G-NIR} = \frac{Green - NIR}{Green + NIR} \quad (5)$$

$$NDFI_{G-MIR} = \frac{Green - MIR}{Green + MIR} \quad (6)$$

$$NDFI_{NIR-MIR} = \frac{NIR - MIR}{NIR + MIR} \quad (7)$$

where *Green*, *NIR* and *MIR* are the intensity values from the green, NIR and MIR wavelengths, respectively.

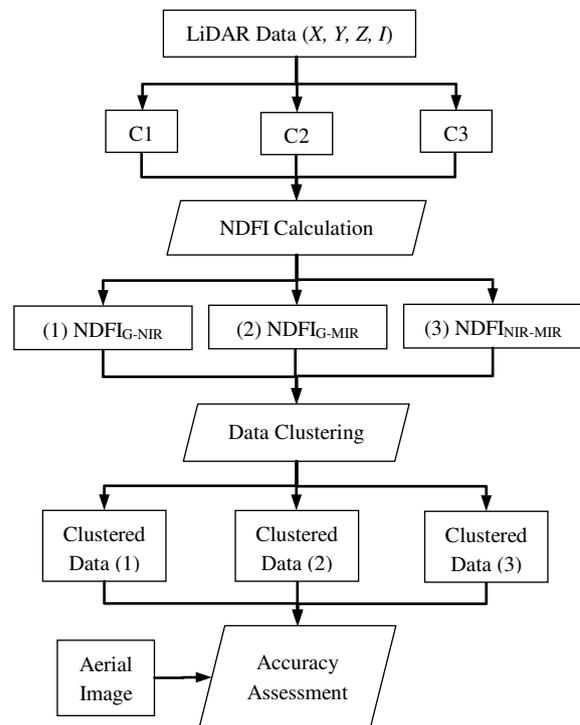


Figure 4. The overall workflow

To be able to calculate different NDFIs, each point from any channel should have a corresponding point from other channels. Although, Optech Titan operates simultaneously at the three wavelengths, it acquires LiDAR points in the three channels at different angles. Consequently, points from different channels do not coincide at the same GPS time. Another method could be used to find corresponding points, called “3D spatial join”, where a search of the nearest neighbours for each point within a sphere of defined radius, using geometric information (x, y, z), is conducted. This method may result in wrong match between points, shown in Figure 5. For example, case (1) indicates the perfect point matching from two channels, C2 and C3. In case (2), a point from C2 could be matched twice with two different points from C3, as this point is the nearest neighbour to both points. Case (3) shows two possible neighbour points from C2 which have the same distance to a point from C3, thus two points could be wrongly matched. Case (4) indicates that no

neighbour points from C2 to a point from C3 within a defined sphere. Therefore, for each channel, a grid was created with 1m cell size and the mean intensity value of each cell was calculated. The grid cell size is selected as 1m in order to guarantee two conditions for each cell; the first is to have sufficient number of points and the second is not to contain any points from different features. Then, the mathematical operation of NDFIs was conducted using grid cells. Based on the point's location, a NDFI value was assigned to each point using bilinear interpolation, where the adjacent cell centres were used in calculation. The whole process is shown in Figure 6.

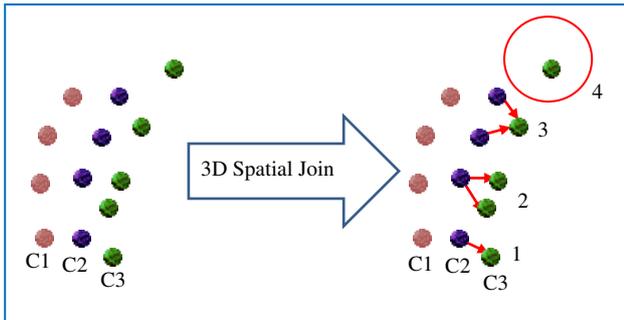


Figure 5. 3D spatial join between C2 and C3 points

A threshold value is required to cluster LiDAR data based on NDFI values, which range from  $-1$  to  $1$ , into two different features. McFeeters (1996) and Xu (2007) set zero as the threshold value for NDWI and NDBI to extract water bodies and built-up areas, respectively, while Chen et al. (2009) adjusted manually a threshold value for NDVI to extract vegetation. The manual adjustment of the threshold could achieve more accurate results among different datasets or sites (Xu, 2006; Lacaux et al., 2007; Li et al., 2013). Therefore, a dynamic threshold adjustment method is required (Ji et al., 2009).

In this study, the Jenks natural breaks optimization method was used to determine the threshold values for discriminating water from land in Site I and vegetation from built-up areas in Site II (Jenks, 1967). This optimization is a data clustering method designed to minimize within-class deviations and thus leads to maximize the between-classes variance ( $\sigma^2$ ), which can be obtained from:

$$\sigma^2 = (M_1 - M)^2 + (M_2 - M)^2 \quad (8)$$

Now, assume that the NDFI points range from  $[a, \dots, b]$ , where  $-1 \leq a < b \leq 1$ . Using the Jenks natural breaks optimization, the ( $M$ ) is first calculated. Then, the points are divided into two classes with ranges  $[a, \dots, t]$  and  $[t, \dots, b]$ , where ( $t$ ) is the threshold value. For each range,  $M_1$  and  $M_2$  are calculated. The optimal ( $t$ ) is obtained by minimizing variation within classes, and intuitively maximizing the between-class variance of the two classes, using the following equation:

$$t = \arg \max_{a \leq t \leq b} \{(M_1 - M)^2 + (M_2 - M)^2\} \quad (9)$$

For land-water mapping of Site I, it is noticed that the water body has high variation in intensity values, especially the intermediate part. A radiometric correction model, developed by (Yan et al., 2012), was applied. The model is based on radar equation and the correction is mainly for the range and the scan angle. Since the high variation of intensity values is at nearly the same range and very small scan angles, the results are insufficient. Therefore, a low pass filter was applied to the intensity data as a pre-processing step before NDFIs calculation, thus making water intensity values more consistent. However, the radiometric correction should be applied in case of processing of overlapped strips (Yan and Shaker, 2015). In addition, the intensity values of water points from the green channel are lower than intensity values from NIR and MIR. Moreover, for each channel, vegetation and water points have a close range of intensity values. As a result, parts of vegetation area and water body could have the same range of NDFIs, thus affecting the classification results of NDFIs for land-water mapping. Therefore, the clustered classes from NDFIs were constrained by the recorded number of returns from different features at the green, NIR, and MIR channels, shown in Table 3.

Feature	Green	NIR/MIR
Built-up areas	single	single
Water	single/double	single
Vegetation	double or more	double or more

Table 3. Recorded returns from different feature in green, NIR, and MIR channels

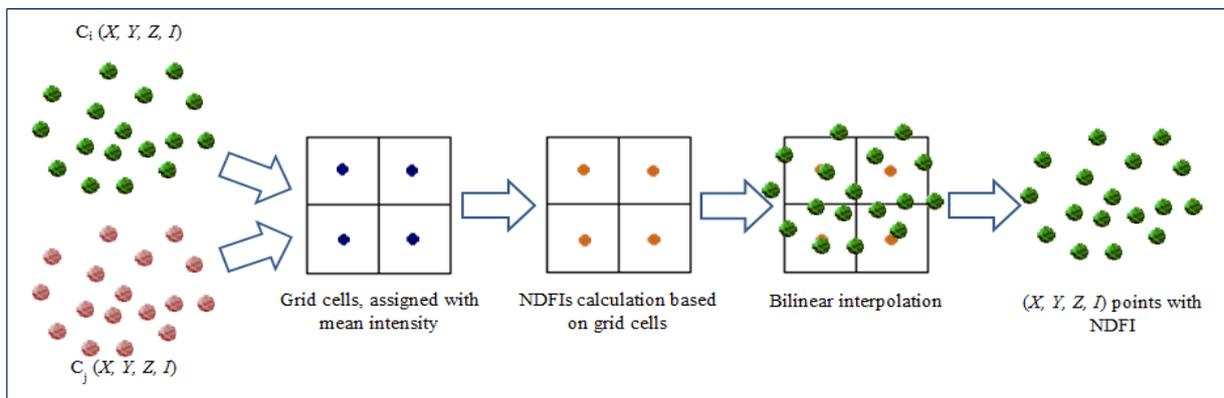


Figure 6. NDFIs calculation for 3D points

#### 4. RESULTS AND DISCUSSIONS

The histogram of each NDFI was first created and studied for both sites. NDFI histograms give an indication of how many classes could be distinguished from the dataset based on number of modes. After that, the Jenks natural breaks optimization method was used to determine the threshold values for each NDFI. The accuracy assessment was conducted using the labeled reference points from the Google image. The confusion matrix was created and the accuracy measures (overall, producer's, and user's accuracies), as well as Kappa statistic were calculated.

##### 4.1 Site I Results

Site I is located at Lake Ontario and covers different land features, such as building, roads, parking lots, trees, grass, and shrubs, as well as water body. Therefore, Site I was tested to extract water body from land features. Figure 7 shows the three histograms, created from the three NDFIs, for Site I.  $NDFI_{G-NIR}$  histogram shows two modes in the histogram indicating two distinguishable classes, while  $NDFI_{G-MIR}$  and  $NDFI_{NIR-MIR}$  histograms show skewed distributions to left and right, respectively, indicating that two classes might be distinguished with difficulty. However, a threshold value was applied to cluster the dataset based on each NDFI into two classes. So that if the  $NDFI_{G-NIR} \leq -0.45$ ,  $NDFI_{G-MIR} \leq -0.41$ , or  $NDFI_{NIR-MIR} > 0.1$ , the cover type is water body; otherwise, it is land.

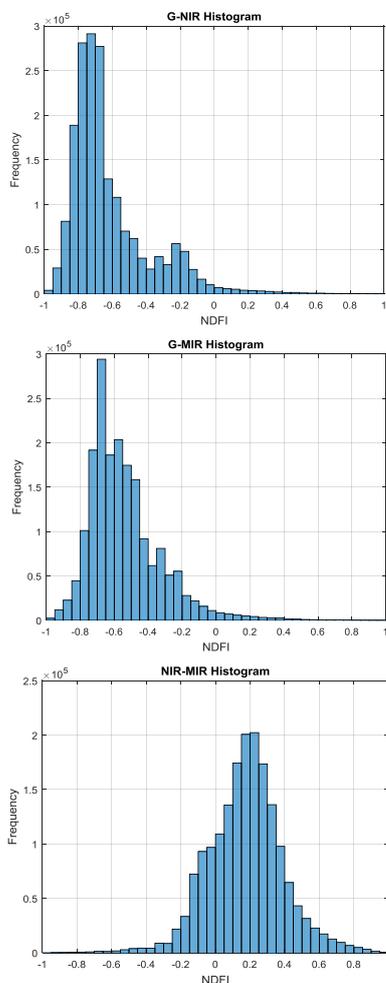


Figure 7. NDFI histograms of Site I

Figure 8 shows the labeled points, as water, land and unclassified, based on  $NDFI_{G-NIR}$  and  $NDFI_{G-MIR}$ , as well as  $NDFI_{NIR-MIR}$ . The results show that most of water points are misclassified as vegetation points. The overall accuracies are 40.31%, 32.19%, and 35.49% using  $NDFI_{G-NIR}$ ,  $NDFI_{G-MIR}$ , and  $NDFI_{NIR-MIR}$ , respectively. The results are not promising for land-water mapping from airborne multispectral LiDAR data using NDFIs only. This could be due to the transmitted signal power of the green wavelength which is lower than the transmitted signal of the NIR and MIR wavelengths. As a result, the intensity values from the green channel are lower than NIR and MIR channels, and hence intensity range values of vegetation and water are very close from each channel. In addition, the water surface has intensity variation, especially in the intermediate part (scan angle near zero).

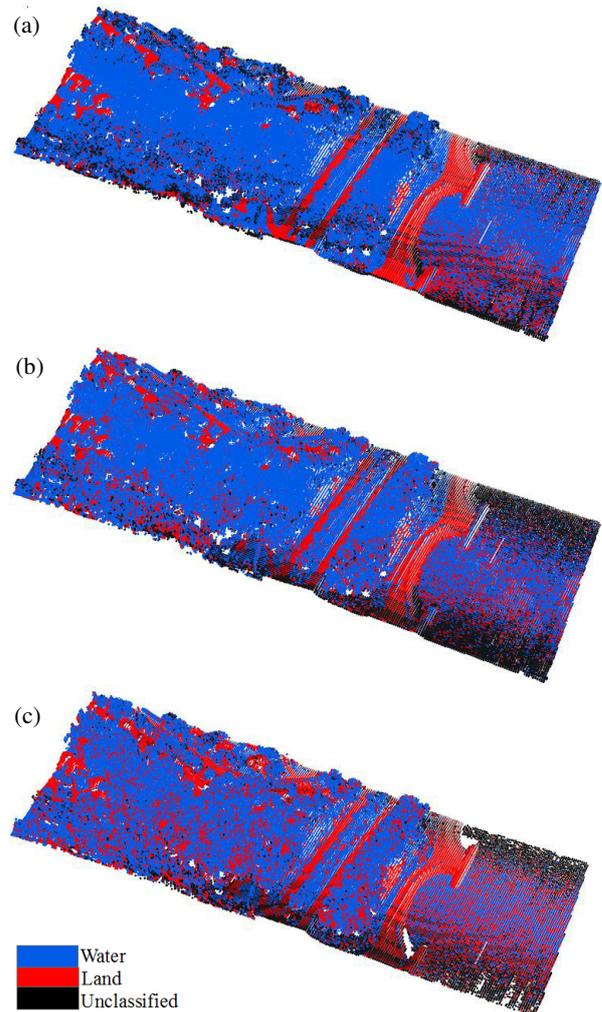


Figure 8. Labeled LiDAR points of Site I from: a)  $NDFI_{G-NIR}$ , b)  $NDFI_{G-MIR}$ , and c)  $NDFI_{NIR-MIR}$

The intensity variation of the water surface was first reduced using a low pass filter. After that, the NDFI calculation was conducted, and a threshold value was applied. The cover type is vegetation and water when the  $NDFI_{G-NIR} \leq -0.51$  or  $NDFI_{G-MIR} \leq -0.53$ ; otherwise, it is built-up area. The output “vegetation and water” class was then constrained by recorded number of returns as shown in Table 3, to extract water points. The cover type is water if the number of returns of the examined point equals 1 or 2 from green channel and 1 from NIR/MIR channel when using  $NDFI_{G-NIR}$  or  $NDFI_{G-MIR}$ , respectively. The  $NDFI_{NIR-MIR}$  is not applicable in this case because it is based on

the NIR and MIR channels, while the green channel should be used to detect water points. Figure 9 shows the labelled points, as water, land and unclassified, based on number of returns added in with the  $NDFI_{G-NIR}/NDFI_{G-MIR}$ . The confusion matrix, overall accuracy, and kappa statistics for the results of the two NDFIs are provided in Table 4 and 5.

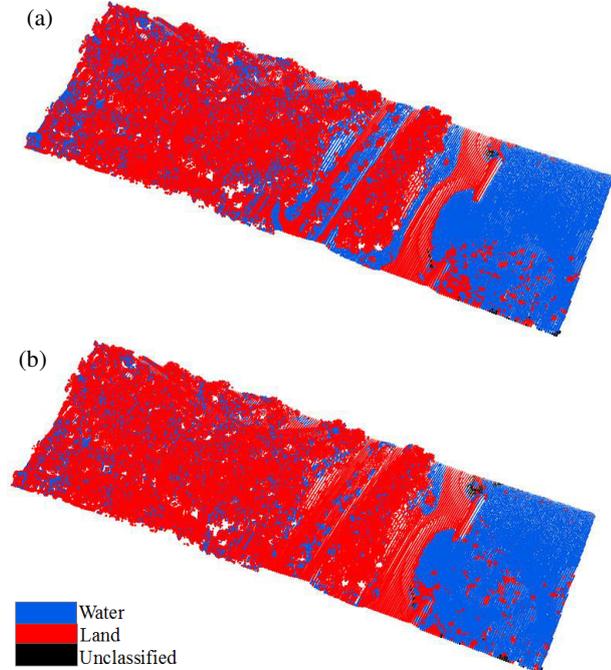


Figure 9. Labelled LiDAR points of Site I from number of returns with a)  $NDFI_{G-NIR}$  / b)  $NDFI_{G-MIR}$

Classification data	Reference data		Total row	User's Acc. (%)
	Water	Land		
Water	61645	3666	65311	94.39
Land	2239	86486	88725	97.48
<b>Total column</b>	63884	90152	154036	
<b>Producer's Acc. (%)</b>	96.50	95.93		

Overall accuracy: 96.17%, overall Kappa statistic: 0.921.

Table 4. Confusion matrix of labelled point based on  $NDFI_{G-NIR}$  and number of returns in Site I

Classification data	Reference data		Total row	User's Acc. (%)
	Water	Land		
Water	61696	2468	64164	96.15
Land	2188	87684	89872	97.57
<b>Total column</b>	63884	90152	154036	
<b>Producer's Acc. (%)</b>	96.58	97.26		

Overall accuracy: 96.98%, overall Kappa statistic: 0.938.

Table 5. Confusion matrix of labelled point based on  $NDFI_{G-MIR}$  and number of returns in Site I

Despite the high overall accuracy, some vegetation areas are misclassified as water points and vice versa. Some low vegetation points (e.g., grass) could not be filtered out. This is because that those points and water points have a close range of intensity values, especially after applying the low pass filter. Also, they have a single return from green, NIR, or MIR

channels. Hence, these vegetation points were misclassified as water points.

#### 4.2 Site II Results

Site II covers an urban area mixed with vegetation (includes: trees, grass, and shrubs) and built-up areas (includes: building, roads, and parking lots). Site II was used to discriminate vegetation from built-up areas. Figure 10 shows the three histograms, created from the three NDFIs, for Site II.  $NDFI_{G-NIR}$  and  $NDFI_{NIR-MIR}$  histograms show two normal distributions, indicating two distinguishable classes, while  $NDFI_{G-MIR}$  histogram shows a skewed distribution to left, indicating that two classes might be distinguished with difficulty. A threshold value was applied to cluster the dataset based on each NDFI into two classes. It is considered that the cover type is vegetation when the  $NDFI_{G-NIR} \leq -0.48$ ,  $NDFI_{G-MIR} \leq -0.42$ , or  $NDFI_{NIR-MIR} > 0.13$ ; otherwise, it is built-up area.

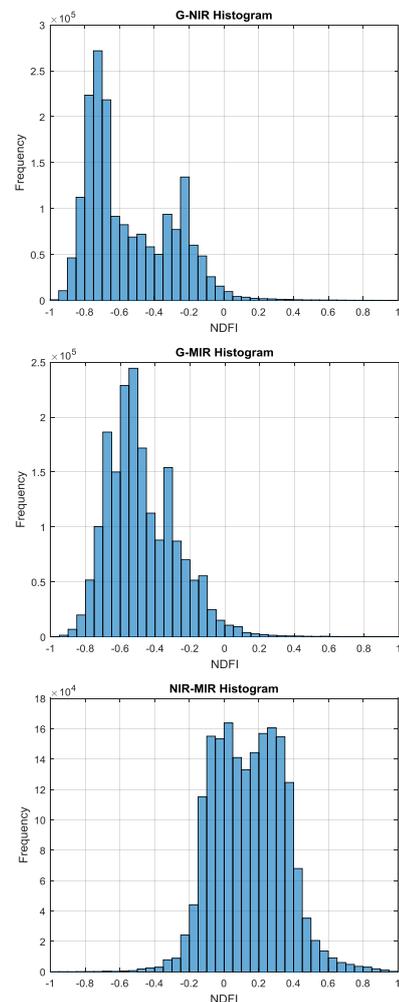


Figure 10. NDFI histograms of Site II

Figure 11 shows the labeled points, as vegetation (V), built-up areas (B) and unclassified (U), based on  $NDFI_{G-NIR}$  and  $NDFI_{G-MIR}$ , as well as  $NDFI_{NIR-MIR}$ . The confusion matrix, overall accuracy, and kappa statistics for the results of the three NDFIs are provided in Table 6, 7 and 8. The  $NDFI_{G-MIR}$  achieved the highest overall accuracy of 92.51%. In particular, the same index,  $NDFI_{G-MIR}$ , achieved the highest vegetation detection rate of 93.17%, while the  $NDFI_{NIR-MIR}$  achieved the highest built-up areas detection rate of 93.79%.

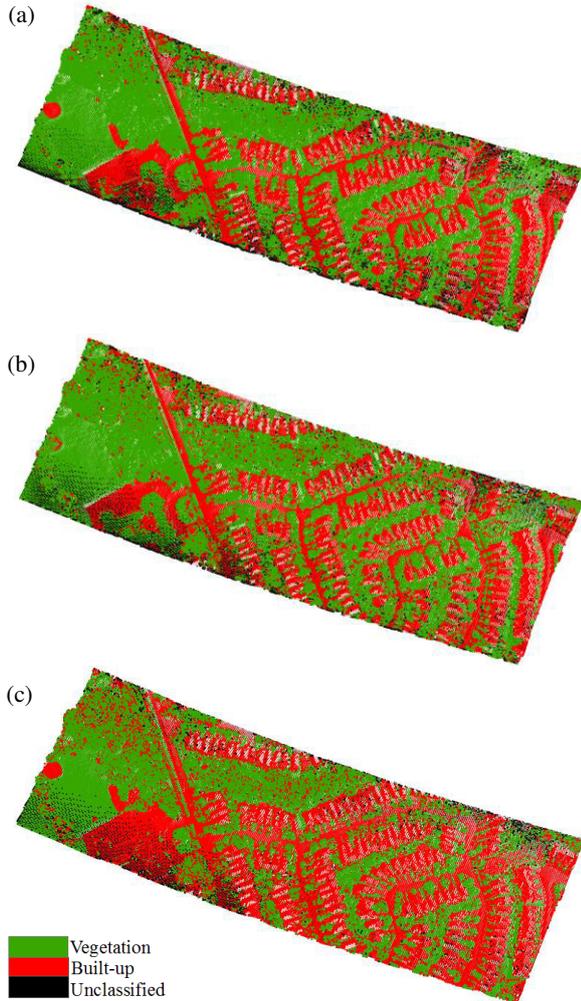


Figure 11. Labelled LiDAR points of Site II from: a)  $NDFI_{G-NIR}$ , b)  $NDFI_{G-MIR}$ , and c)  $NDFI_{NIR-MIR}$

Classification data	Reference data			Total row	User's Acc. (%)
	U	V	B		
U	0	2980	2133	5113	0
V	0	52773	31	52804	99.94
B	0	3016	27835	30851	90.22
<b>Total column</b>	0	58769	29999	88768	
<b>Producer's Acc. (%)</b>		89.80	92.79		

Overall accuracy: 90.81%, overall Kappa statistic: 0.812.

Table 6. Confusion matrix of labelled point based on  $NDFI_{G-NIR}$  in Site II

Classification data	Reference data			Total row	User's Acc. (%)
	U	V	B		
U	0	1163	1942	3105	0
V	0	54756	696	55452	98.74
B	0	2850	27361	30211	90.57
<b>Total column</b>	0	58769	29999	88768	
<b>Producer's Acc. (%)</b>		93.17	91.21		

Overall accuracy: 92.51%, overall Kappa statistic: 0.841

Table 7. Confusion matrix of labelled point based on  $NDFI_{G-MIR}$  in Site II

Classification data	Reference data			Total row	User's Acc. (%)
	U	B	V		
U	0	1484	1061	2545	0
B	0	28091	12533	40624	69.15
V	0	375	50950	51325	99.27
<b>Total column</b>	0	29950	64544	94494	
<b>Producer's Acc. (%)</b>		93.79	78.94		

Overall accuracy: 83.65%, overall Kappa statistic: 0.668.

Table 8. Confusion matrix of labelled point based on  $NDFI_{NIR-MIR}$  in Site II

## 5. CONCLUSIONS AND FUTURE WORK

This paper discusses the use of airborne multispectral LiDAR data for land-water, vegetation, and built-up area mapping. Three NDFIs namely,  $NDFI_{G-NIR}$ ,  $NDFI_{G-MIR}$ , and  $NDFI_{NIR-MIR}$  were evaluated using intensity values from green, NIR, and MIR wavelengths of the new airborne sensor "Optech Titan". The Jenks natural breaks optimization was used to determine the threshold values of each NDFI, in order to avoid the manual adjustment of threshold values and cluster the 3D point data. The results show low performance of using NDFIs only in land-water mapping. The low performance may be due to the variation of the transmitted signal power at different wavelength. Hence, the recorded intensity for an object (e.g., water) is high at NIR and MIR wavelengths compared to green wavelength. Consequently, water points are misclassified as vegetation points. The land-water mapping results were significantly improved by using two steps before and after NDFIs calculation. A low pass filter was applied to intensity data as a pre-processing step before NDFIs calculation to reduce the intensity variation in the water surface. The output "vegetation and water" class from clustering was constrained using recorded number of returns from different channels. On the other hand, the direct use of the three NDFIs succeeded to discriminate vegetation from built-up areas. Based on the results achieved, we suggest using the  $NDFI_{G-MIR}$  and  $NDFI_{NIR-MIR}$  for vegetation and built-up areas extraction, respectively, from airborne multispectral LiDAR data. The discrimination of roads from buildings and low from high vegetation will be further investigated by incorporating the LiDAR height data. Also, other factors will be introduced to distinguish water from land features.

## ACKNOWLEDGEMENTS

This research work is supported by the Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Ontario Trillium Scholarship (OTS). The authors also would like to thank Teledyne Optech for providing the LiDAR data from the world's first multispectral LiDAR system, the Optech Titan.

## REFERENCES

Chen, Y., Su, W., Li, J., Sun, Z., 2009. Hierarchical object oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research*, 43(7), pp. 1101-1110.

- Gao, B. C., 1996. NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3), pp. 257-266.
- Guan, H., Ji, Z., Zhong, L., Li, J., Ren, Q., 2013. Partially supervised hierarchical classification for urban features from lidar data with aerial imagery. *International journal of remote sensing*, 34(1), pp. 190-210.
- Hartfield, K. A., Landau, K. I., Van Leeuwen, W. J., 2011. Fusion of high resolution aerial multispectral and LiDAR data: Land cover in the context of urban mosquito habitat. *Remote Sensing*, 3(11), pp. 2364-2383.
- Huang, M. J., Shyue, S. W., Lee, L. H., Kao, C. C., 2008. A knowledge-based approach to urban feature classification using aerial imagery with lidar data. *Photogrammetric Engineering & Remote Sensing*, 74(12), pp. 1473-1485.
- Jenks, G. F., 1967. *The data model concept in statistical mapping*. International yearbook of cartography, 7(1), pp. 186-190.
- Ji, L., Zhang, L., Wylie, B., 2009. Analysis of dynamic thresholds for the normalized difference water index. *Photogrammetric Engineering & Remote Sensing*, 75(11), pp. 1307-1317.
- Lacaux, J. P., Tourre, Y. M., Vignolles, C., Ndione, J. A., Lafaye, M., 2007. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote Sensing of Environment*, 106(1), pp. 66-74.
- Li, W., Du, Z., Ling, F., Zhou, D., Wang, H., Gui, Y., Sun, B., Zhang, X., 2013. A comparison of land surface water mapping using the normalized difference water index from TM, ETM+ and ALI. *Remote Sensing*, 5(11), pp. 5530-5549.
- MacFaden, S. W., O'Neil-Dunne, J. P., Royar, A. R., Lu, J. W., Rundle, A. G., 2012. High-resolution tree canopy mapping for New York City using LIDAR and object-based image analysis. *Journal of Applied Remote Sensing*, 6(1), pp. (063567)1-23.
- McFeeters, S. K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, 17(7), pp. 1425-1432.
- Puttonen, E., Hakala, T., Nevalainen, O., Kaasalainen, S., Krooks, A., Karjalainen, M., Anttila, K., 2015. Artificial target detection with a hyperspectral LiDAR over 26-h measurement. *Optical Engineering*, 54(1), pp. 013105-013105.
- Rogers, A. S., Kearney, M. S., 2004. Reducing signature variability in unmixing coastal marsh Thematic Mapper scenes using spectral indices. *International Journal of Remote Sensing*, 25(12), pp. 2317-2335.
- Rouse Jr, J., Haas, R. H., Schell, J. A., Deering, D. W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication*, 351, p. 309.
- Shuo, S., Shalei, S., Wei, G., Lin, D., Bo, Z., Xin, H., 2015. Improving Backscatter Intensity Calibration for Multispectral LiDAR. *Geoscience and Remote Sensing Letters, IEEE*, 12(7), pp. 1421-1425.
- Thanapura, P., Helder, D., Burckhard, S., Warmath, E., O'Neill, M., Galster, D., 2007. Mapping urban land cover using Quickbird NDVI and GIS spatial modeling for runoff coefficient determination. *Photogrammetric Engineering & Remote Sensing*, 73(1), pp. 57-65.
- Wei, G., Shalei, S., Bo, Z., Shuo, S., Faquan, L., Xuewu, C., 2012. Multi-wavelength canopy LiDAR for remote sensing of vegetation: Design and system performance. *ISPRS Journal of Photogrammetry and Remote Sensing*, 69, pp. 1-9.
- Wichmann, V., Bremer, M., Lindenberger, J., Rutzinger, M., Georges, C., Petrini-Monteferrri, F., 2015. Evaluating the Potential of Multispectral Airborne LIDAR for Topographic Mapping and Land Cover Classification. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 113-119.
- Woodhouse, I. H., Nichol, C., Sinclair, P., Jack, J., Morsdorf, F., Malthus, T. J., Patenaude, G., 2011. A multispectral canopy LiDAR demonstrator project. *Geoscience and Remote Sensing Letters, IEEE*, 8(5), pp. 839-843.
- Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14), pp. 3025-3033.
- Xu, H., 2007. Extraction of urban built-up land features from Landsat imagery using a thematic-oriented index combination technique. *Photogrammetric Engineering & Remote Sensing*, 73(12), pp. 1381-1391.
- Yan, W. Y., Shaker, A., Habib, A., & Kersting, A. P., 2012. Improving classification accuracy of airborne LiDAR intensity data by geometric calibration and radiometric correction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 35-44.
- Yan, W. Y., Shaker, A., 2015. Radiometric normalization of overlapping LiDAR intensity data for reduction of striping noise. *International Journal of Digital Earth*, 1-13.
- Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), pp. 583-594.
- Zhou, W., Huang, G., Troy, A., Cadenasso, M. L., 2009. Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study. *Remote Sensing of Environment*, 113(8), pp. 1769-1777.