USING LIDAR SURVEY FOR LAND USE CLASSIFICATION

C Ryan, Catchment Simulation Solutions, Sydney, NSW

Special thanks to:
J Molteno, Blacktown City Council, Sydney, NSW
J Ruszczyk, Warringah Council, Sydney, NSW

Abstract

Light Detection and Ranging (commonly termed LiDAR or ALS) survey is commonly used in flooding investigations to define the Digital Elevation Model (DEM) used by many hydrologic and hydraulic modelling software. Australian Councils often make a considerable financial investment in obtaining this type of survey and it is important that this data is utilised to its fullest potential to produce the greatest benefit for Council.

At a cursory examination, LiDAR seems to contain simply elevation data associated with each point. Some survey providers separate ground from non-ground points which can inform building footprints. However, it is not commonly known that significant additional information can be extracted from the raw data which includes intensity, multiple returns and other useful parameters. The author contends that the full range of data can be analysed with remote sensing based approaches to create other useful datasets besides elevation.

In this paper, we present two case studies where the full range of LiDAR information is used in conjunction with automated analysis of aerial photography to differentiate several land use categories and estimate impervious proportions. This is done with a view to determining Manning's 'n' material types for TUFLOW hydraulic modelling but could also be useful for deriving loss rates and impervious proportions in a hydrologic model amongst other uses. The paper discusses the methods used to implement the analysis as well as the benefits and challenges involved.

Introduction

Light Detection and Ranging (LiDAR) survey is playing an increasingly important role in flooding investigations. It is well known that LiDAR survey can provide highly detailed and accurate elevation data to support our hydrologic and hydraulic models. However, it is less well known that significant additional data is routinely recorded during LiDAR survey including intensity, multiple returns and other parameters. This data can be used to infer additional information regarding the materials and surfaces on the ground. In fact, since the release of the LAS 1.2 format specifications in September 2008, it is now mandatory for survey providers to complete some classification of points into categories including ground, vegetation, buildings, water and others. These classifications can also be useful to inform our hydrologic and hydraulic models.

This paper will also argue that by considering other commonly available data such as high resolution aerial photographs in conjunction with LiDAR, further classification can be to undertaken to breakup ground points into impervious (such as roads and car parks) and non-impervious (such as grass and dirt) surfaces. Such classification can add significant value to our models since impervious areas and loss rates are often key project inputs.

The evolution of the LiDAR LAS format standards and resulting classification regimes will be described in the following section. Following this, methods to combine LiDAR with aerial photography to improve classification will be discussed. Two case studies will also be presented to demonstrate the concepts.

Evolution of LiDAR Data Standards

The LASer (LAS) file format has become an industry standard public file format for point cloud data such as LiDAR. The format is managed by the American Society for Photogrammetry and Remote Sensing (ASPRS). The LAS 1.0 format was released in May, 2003 and since then, there have been 4 new version of the standard that have been approved by the ASPRS board culminating in the most recent standard, LAS 1.4 in November 2011. As the standards have evolved the requirements for survey providers to store additional information and incorporate classification have also changed.

LAS 1.2 (September 2008) was the first standard to require survey providers to include classification in the LAS files. This format supported 4 permissible Point Data Record Formats (PDRF). The basic PDRF is shown below in Table 1. The other PDRFs in LAS 1.2 allowed for the addition of Global Positioning System (GPS) time and Red-Green-Blue (RGB) colour records.

Item	Format	Size	Required
X	long	4 bytes	*
Y	long	4 bytes	*
Z	long	4 bytes	*
Intensity	unsigned short	2 bytes	
Return Number	3 bits (bits 0, 1, 2)	3 bits	*
Number of Returns (given pulse)	3 bits (bits 3, 4, 5)	3 bits	*
Scan Direction Flag	1 bit (bit 6)	1 bit	*
Edge of Flight Line	1 bit (bit 7)	1 bit	*
Classification	unsigned char	1 byte	*
Scan Angle Rank (-90 to +90) – Left side	char	1 byte	*
User Data	unsigned char	1 byte	
Point Source ID	unsigned short	2 bytes	*

Table 1: Point Data Format 0 LAS Format 1.2 (September 2008)

The classification required for LAS 1.2 is shown below in Table 2.

ASPRS Standard LIDAR Point Classes

Classification Value (bits 0:4)	***Meaning***
0	Created, never classified
1	Unclassified[1]
2	Ground
3	Low Vegetation
4	Medium Vegetation
5	High Vegetation
6	Building
7	Low Point (noise)
8	Model Key-point (mass point)
9	Water
10	*Reserved for ASPRS Definition*
11	*Reserved for ASPRS Definition*
12	Overlap Points[2]
13-31	*Reserved for ASPRS Definition*

Table 2: Standard LIDAR Classification (mandatory since LAS 1.2 format)

LAS format 1.4 now supports 11 different types of PDRF. Although a similar PDRF to that shown in Table 1 is still available, survey providers are encouraged to use a more detailed PDRF such as the one shown below in Table 3 (PDRF #7).

Table 18: Point Data Record Format 7

Item	Format	Size	Required
X	long	4 bytes	*
Y	long	4 bytes	*
Z	long	4 bytes	*
Intensity	unsigned short	2 bytes	
Return Number	4 bits (bits 0, 1, 2, 3)	4 bits	*
Number of Returns (given pulse)	4 bits (bits 4, 5, 6, 7)	4 bits	*
Classification Flags	4 bits (bits 0 – 3)	4 bits	
Scanner Channel	2 bits (4-5)	2 bits	*
Scan Direction Flag	1 bit (bit 6)	1 bit	*
Edge of Flight Line	1 bit (bit 7)	1 bit	*
Classification	unsigned char	1 byte	*
User Data	unsigned char	1 byte	
Scan Angle	short	2 bytes	*
Point Source ID	unsigned short	2 bytes	*
GPS Time	double	8 bytes	*
Red	unsigned short	2 bytes	*
Green	unsigned short	2 bytes	*
Blue	unsigned short	2 bytes	*

Table 3: Point Data Record Format 7 LAS Format 1.4

As shown above, additional data is required for these more detailed PDRFs. Recognising the potential for these fields to aid in more detailed classification, the standards have a different classification standard for these PDRFs as shown in Table 4.

Table 17: ASPRS Standard LIDAR Point Classes (Point Data Record Formats 6-10)

Classification Value	*Meaning*
0	Created, never classified
1	Unclassified[3]
2	Ground
3	Low Vegetation
4	Medium Vegetation
5	High Vegetation
6	Building
7	Low Point (noise)
8	Reserved
9	Water
10	Rail
11	Road Surface
12	Reserved
13	Wire – Guard (Shield)
14	Wire – Conductor (Phase)
15	Transmission Tower
16	Wire-structure Connector (e.g. Insulator)
17	Bridge Deck
18	High Noise
19-63	Reserved
64-255	User definable

Table 4: Standard LIDAR Classification for preferred Point Data Formats in LAS 1.4 format

These more detailed classifications include road surfaces, bridge decks and other classifications that would be extremely useful for flooding investigations. However, the author has not yet found any LAS files conforming to these standards so their effectiveness remains to be seen.

The LAS format specifications appear to be continually evolving to ensure the types of data that may be collected as LiDAR evolves can be stored in a standardised manner. However, it remains to be seen how quickly the industry will evolve to utilise the more detailed formats and PDRFs available. Furthermore, due to the cost of LiDAR it will be necessary to continue to use old LAS formats with no or limited classification for several years into the future. As such, there is significant potential for consultants and practitioners to add value to LiDAR survey by post processing for improved land use classification and impervious area estimation to support our flood models.

Combination with Aerial Photography

As noted in the preceding section, all the LAS formats discussed have PDRFs that can support RGB colour values captured during LiDAR survey. Even if RGB values are not being captured, high resolution aerial photography is normally available for use in our flooding investigations. As such, it becomes attractive to investigate ways that aerial photography can be used to improve automated land use classification in conjunction with LiDAR data.

A common method for doing this is based on the Normalised Vegetation Difference Index (NVDI) algorithm which aims to combine near infra-red (NIR) band values with the red band (VIS) in the visible spectrum to create an index with the following formula.

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{VIS})}{(\mathrm{NIR} + \mathrm{VIS})}$$

Equation 1: Normalized Difference Vegetation Index (NDVI)

This index is commonly applied to determine if the target contains green vegetation or not and thus can be applied to attempt to differentiate impervious ground points from other vegetated surfaces such as grass (Kriegler, FJ et al 1969).

Unfortunately, neither LiDAR nor aerial photographs contain proper near infra-red data. However, the intensity field of LiDAR survey is a pseudo measure of infra-red and thus the NVDI algorithm can still be applied (Rottensteiner et al, 2003). We will refer to this as a pseudo-NVDI.

There are also other indexes that can be created from combinations of the aerial photography visible spectrum data and the LiDAR data including the pseudo-green-NVDI (using visible green instead of visible red band) and the Green-Red Vegetation Index (using visible green and visible red instead of near infra-red) Motohka et al (2010). In certain circumstances, these indexes may also be used to aid in further classification. However, for the purposes of this investigation, only the pseudo NVDI has been used.

Two cases studies are presented in the following sections which demonstrate how a pseudo-NVDI image can be used to create new or augment existing LiDAR classifications.

Case Study 1 – Eastern Creek Catchment – Hydraulic Assessment

Catchment Simulation Solutions (CSS) was engaged to complete the Eastern Creek Catchment – Hydraulic Assessment in 2012. As part of this project a TUFLOW hydraulic model was to be created and CSS proposed to utilise Council's LiDAR LAS data to undertake classification and inform material types for the model.

After examining the LiDAR information, it became apparent that no existing classifications were included in the LAS files. As such, a complete workflow was constructed to classify ground, buildings and tree canopy using the LAStools software suite developed by RapidLasso (http://rapidlasso.com/). This workflow included:

1. Lasground.exe to classify ground vs non-ground
2. Lasheight.exe to calculate height of non-ground points
3. Las2dem.exe to create grids for Digital Elevation Model (DEM), Digital Surface Model (DSM) and Intensity values
4. Lasgrid.exe to create grid of the number of returns
5. Lasclassify.exe to classify buildings vs tree canopy

6. Lasgrid.exe to create raster (ie., grid) representation of buildings
7. Lasboundary.exe to create vector (ie., polygon) representation of buildings

This workflow can effectively differentiate between buildings, tree canopy and ground points as well as creating elevation surfaces that include or exclude buildings and trees. However, it does not differentiate between impervious versus non-impervious ground points. To aid in this classification we employed the pseudo-NVDI algorithm described in the preceding section. This algorithm was applied using the open source Geospatial Data Abstraction Library (GDAL - http://www.gdal.org/) and some internally developed tools.

After applying the pseudo-NVDI algorithm it became clear that this index could differentiate between road and vegetated ground areas with reasonable accuracy. A classification schema tailored to the project's needs was applied as follows:

- Buildings classified by LAStools workflow were retained
- If height > 2m and number of returns > 1 then classify as tree canopy
- If intensity < 8 or (intensity < 30 and NVDI > 0) then classify as water
- If NVDI < -0.5 then classify as roads
- If NVDI between -0.5 and -0.1 then classify as low vegetation (low greenness)
- If NVDI > -0.1 then classify as low vegetation (high greenness)

A classification of a sample area from the Eastern Creek catchment is shown in Figure 1.

Figure 1: Classification of Eastern Creek Area

During the classification exercise, a number of issues and misclassifications were observed. These included:

- Bridges and road abutments occasionally classified as buildings
- General difficulties picking up small buildings.
- Shadows affecting the pseudo-NVDI algorithm causing misclassification of roads and vegetation in the shade.
- Overhanding tree canopies making classification of underlying material type difficult.

In subsequent discussions with the LAStools developer, it was suggested that LiDAR point density is critical for good building extraction. 2 points per square metre is recommended and the Eastern Creek LiDAR data has approximately 0.8 points per square metre.

Despite the observed examples of misclassification, the classification was considered successful. A small amount of manual modification was required to create a classification dataset with sufficient resolution and accuracy to support hydraulic flood modelling. The alternative of manual digitising of material types would have been prohibitively time consuming considering the size of the project area.

Case Study 2 – Narrabeen & Dee Why Land Use and Impervious Breakdown

CSS was engaged to contribute to the 'Stormwater Models for Dee Why and Narrabeen Lagoon Catchments' project for Warringah Council in conjunction with Catchment Research Pty Ltd. One of the key deliverables was land use mapping and estimation of impervious proportions for the catchments. LAS data was available from Council.

When the LAS data was examined, it became evident that the survey provider had supplied some classifications in accordance with the classification schema in Table 2. These classifications are visualised for a sample area in Figure 2.

.

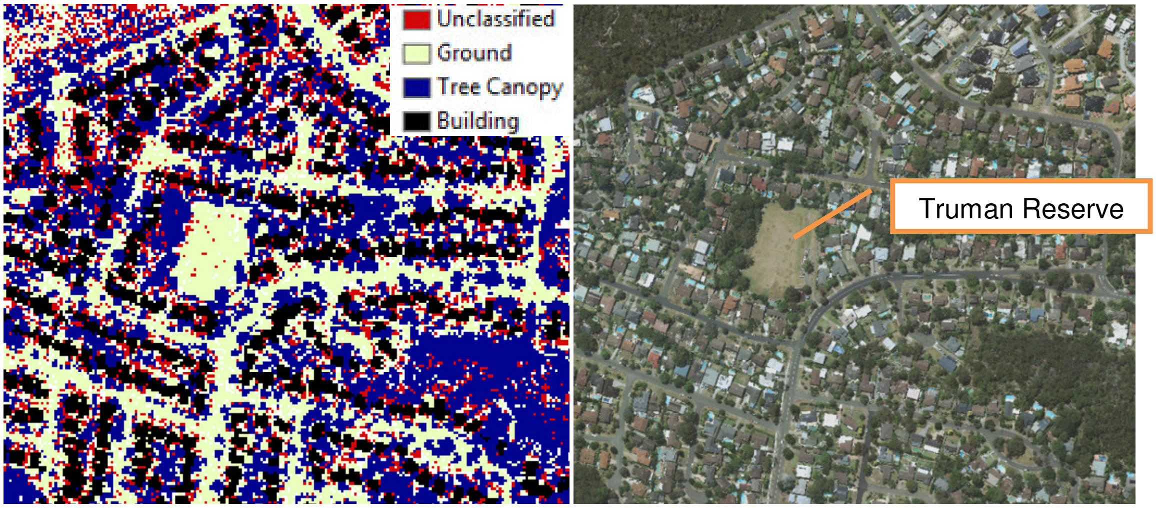

Figure 2: Provider Classifications for Sample Area (Truman Reserve, Cromer)

As shown in Figure 2, these classifications mainly differentiated between building, tree canopy and ground. There were no results for the water, high vegetation or low vegetation categories. More importantly for our purposes, the classifications did not distinguish pervious versus impervious ground points as can be seen when looking at the Truman Reserve which has the same classification as the roads.

In order to try and improve these classifications, we utilised the aerial photographs to create a pseudo-NVDI index as described in the prior sections. The NVDI image for the sample area is shown in Figure 3.

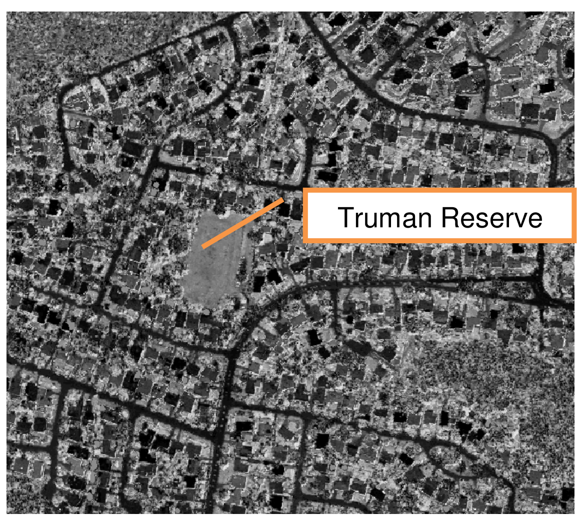

Figure 3: NVDI Image for Sample Area (refer Figure 2)

It can be seen in Figure 3 that there is a clear difference in the ‘colour’ of the NVDI for the impervious surfaces (roads and car parks) versus Truman Reserve and other grassed areas.

A classification schema was then developed that adopted the provider's classifications for buildings and tree canopy. However, we devised our own classification routine for the remaining categories of interest. The classification rules were:

- Provider classification of building was preserved
- Provider classification of tree canopy was preserved
- A pseudo-NVDI index was created from aerial photography and rasterised LiDAR intensity.
- If intensity was 0 or non-defined, then classify as water
- If NVDI was less than -0.65 then classified as dark sealed roads
- If NVDI was between -0.65 to -0.5 then classify as lighter coloured impervious areas (such as concrete)
- If NVDI greater than -0.5 then classify as pervious low vegetation (such as grass)

An example of the classification and the aerial photograph for comparison is shown in Figure 4.

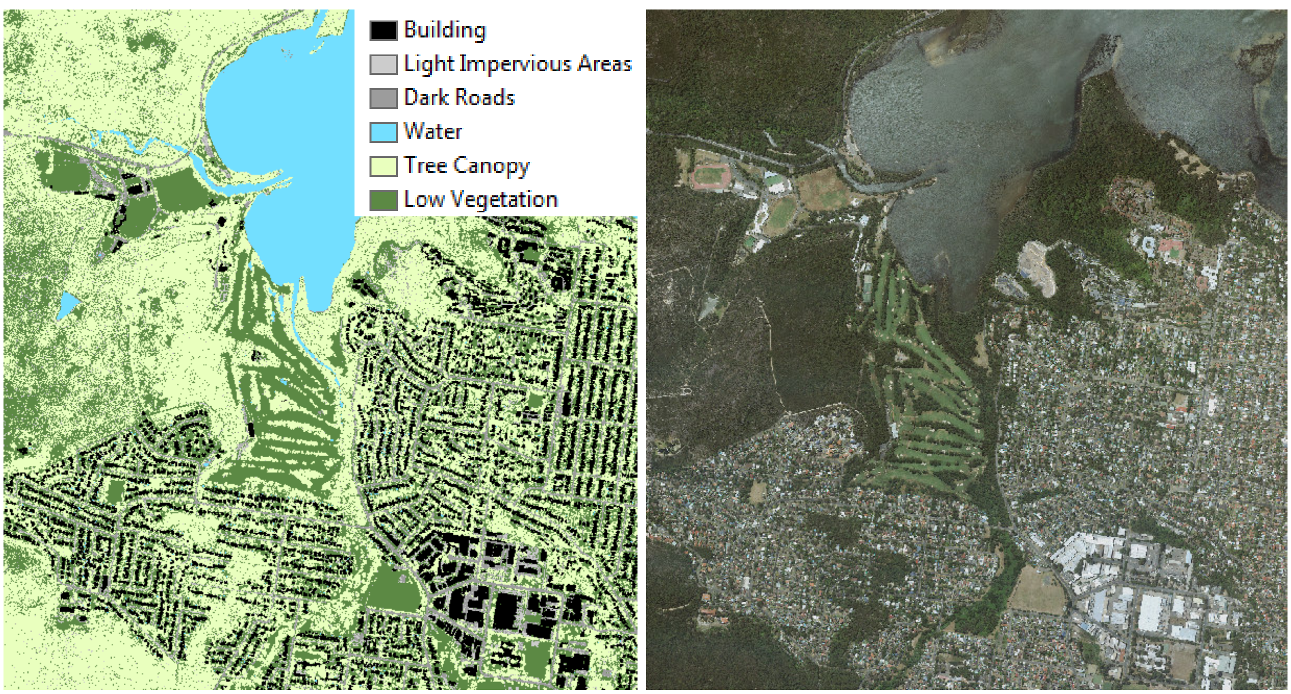

Figure 4: Classification Example for Dee Why and Narrabeen Lagoon Catchments

The classification was successful although there were some issues worth noting. Large parts of the area had significant overhanging tree canopy areas and classification of roads underneath these trees is difficult. As a result, it is likely that roads are under-classified in these areas. However, misclassifications and their consequences need to be viewed in the context of the end use of the data. From a hydrologic perspective, the tree canopy is quite relevant to the rainfall loss rates regardless of the underlying ground cover whereas from a hydraulic flood modelling perspective, we are much more interested in the underlying material type.

Conclusions

This paper has aimed to demonstrate that LiDAR survey can be useful in flooding investigations for land use and imperviousness classification in addition to creation of elevation models. While the industry and LAS standard specifications are moving towards automated classifications of LiDAR, there is still significant opportunity for consultants and practitioners to improve classification, particularly when utilising other data sources such as high resolution aerial photographs.

Two case studies were presented to demonstrate these concepts. In the first case, no classifications were included from the survey provider and a complete workflow was created to classify the land uses. In the second case study, the survey provider had undertaken some classification but this was improved and expanded by using some simple remote sensing techniques.

The techniques employed have demonstrated that many useful datasets besides simple elevation data can be derived from good quality LiDAR survey data. Other uses cases might include calculations of tree canopy extent and related data such as bushfire fuel loadings. Repeated LiDAR survey over time could also be valuable for many uses including identification of unauthorised tree removal and mapping of development rates.

Recommendations to Councils when Acquiring LiDAR Survey

- Find out which LAS format specification will be used and which Point Data Record Format will be adopted.
- Ensure the full dataset will be delivered in LAS format.
- Utilise the free open source LASzip tool to cut data storage by 80-93% without losing any data.
- Ensure a sufficient point density to ensure accurate building extraction (2 points per square metre or more).
- Investigate and consider costs associated with capturing RGB values for LiDAR points.
- Investigate and consider costs associated with capturing full wave form data.

References

Kriegler, F.J., Malila, W.A., Nalepka, R.F., and Richardson, W. (1969) 'Preprocessing transformations and their effects on multispectral recognition.' Proceedings of the Sixth International Symposium on Remote Sensing of Environment, p. 97-131.

Takeshi Motohka, Kenlo Nishida Nasahara, Hiroyuki Oguma and Satoshi Tsuchida (2010), Applicability of Green-Red Vegetation Index for Remote Sensing of Vegetation Phenology. *Remote Sensing, 2010, 2, 2369-2387*

Franz Rottensteiner , John Trinder , Simon Clode , Kurt Kubik (2003), Detection of buildings and roof segments by combining LIDAR data and multispectral images. *Image and Vision Computing,* New Zealand

American Society for Photogrammetry and Remote Sensing, ASPRS LAS 1.0 Format Specification (2003)

American Society for Photogrammetry and Remote Sensing, ASPRS LAS 1.2 Format Specification (2008)

American Society for Photogrammetry and Remote Sensing, ASPRS LAS 1.4 Format Specification (2011)