Statistical analysis of airborne LiDAR data for forest classification in the Strzelecki Ranges, Victoria, Australia

Z. Zhang a, b, X. Liu b, J. Peterson a and W. Wright c

a Centre for GIS, School of Geography and Environmental Science
Monash University, Clayton, Victoria 3800, Australia

b Australian Centre for Sustainable Catchments, and Faculty of Engineering and Surveying
University of Southern Queensland, Toowoomba, Queensland 4350, Australia

c School of Applied Sciences and Engineering
Monash University, Churchill, Victoria 3842, Australia
Email: zhenyu.zhang@usq.edu.au

Abstract: Although remotely sensed data have been widely explored for forest applications, passive remote sensing techniques are limited in their ability to capture forest structural complexity, particularly in uneven-aged, mixed species forests with multiple canopy layers. Generally, these techniques are only able to provide information on horizontal (two-dimensional) forest extent. The vertical forest structure (or the interior of the canopy and understorey vegetation) cannot be mapped using these passive remote sensing techniques. Fortunately, it has been shown that active remote sensing techniques via airborne LiDAR (light detection and ranging) with capability of canopy penetration yields such high density sampling that detailed description of the forest structure in three-dimensions can be obtained. Accordingly, much interest is attached to exploring the application of this approach for identifying the distribution of designated vegetation communities. However, the suitability of LiDAR data for the classification of forests with complex structures, particularly for cool temperate rainforest and neighbouring uneven-aged mixed forests in a severely disturbed landscape has hitherto remained untested.

This study applied airborne LiDAR data for the classification of cool temperate rainforest dominated by Myrtle Beech (Nothofagus cunninghamii) and adjacent forests including naturally regenerated Mountain Ash (Eucalyptus regnans), mixed forest consisting of overstorey Mountain Ash and understorey Myrtle Beech, Silver Wattle (Acacia dealbata), and hardwood plantation dominated by Shining Gum (Eucalyptus nitens) in the Strzelecki Ranges, Victoria, Australia. LiDAR data were extracted within each of the forest plots. Non-ground laser returns were used to generate forest height profiles for the analysis of the spatial distribution of vertical forest structure for the plots dominated by different forest types. The $k$-means clustering algorithm was performed on each of the plots to stratify the vertical forest structure into three layers, representing the overstorey, mid-storey and lower storey of the plot-level forests. Variables were then calculated from the LiDAR data based on the three-layered structure for each plot. The statistical analyses, which included one-way ANOVA (analysis of variance) and the post hoc tests, identified effective variables for forest type classifications. Linear discriminant analysis with cross-validation was carried out to classify the forest types and assess the classification accuracy using error matrices. This study demonstrated the applicability of airborne LiDAR for the classification of the Australian cool temperate rainforest and adjacent forests in the study area.

Keywords: LiDAR, cool temperate rainforest, forest classification, statistical analysis, Strzelecki Ranges
1. INTRODUCTION

Forest type classification is part of decision support for sustainable forest management and native forest conservation (Koch et al., 2006; Brandtberg, 2007). The traditional methods for forest type classification were based either on an interpretation of large-scale aerial photographs or a field inventory work. These methods are labour intensive and time consuming (Franklin, 2001; Persson et al., 2002; Kim et al., 2009). Although remotely sensed data have been widely explored for forest applications, passive remote sensing techniques still fall short of capturing three-dimensional forest structures, particularly in uneven-aged, mixed species forests with multiple canopy layers (Lovell et al., 2003). Fortunately, it has been shown that active remote sensing via airborne LiDAR (light detection and ranging) with capability of canopy penetration yields such high density sampling from the top and interior of the canopy and understory vegetation that detailed description of the forest structure in three-dimensions can be obtained (Popescu et al., 2002; Evans et al., 2006). Accordingly, much interest is attached to exploring the application of airborne LiDAR data for identifying the distribution of designated vegetation communities.

There has been increasing interest in the application of airborne LiDAR for the analysis for forest structures and the classification of forest type over the last decade (Popescu et al., 2002; Zimble et al., 2003; Goodwin et al., 2006; Ørka et al., 2009; Suratno et al., 2009; Vauhkonen et al., 2009; Korpela et al., 2010). The vertical distribution and configuration of the forest components are forest type dependent (Brandtberg, 2007). The structural differences will affect the distribution of the laser returns from the forests (Ørka et al., 2009). Therefore, variables that characterise the height distribution and density of the canopy can be generated from the LiDAR data and complementary field data, for tree species identification and classification (Næsset, 2001; Heurich and Thoma, 2008; Korpela et al., 2010). The applications of LiDAR-derived variables for forest classification have been attempted for coniferous forests (Suratno et al., 2009), deciduous forest (Brandtberg, 2007), and mixed coniferous, deciduous and other forests (Moffiet et al., 2005; Heurich and Thoma, 2008; Holmgren et al., 2008; Ørka et al., 2009; Vauhkonen et al., 2009; Korpela et al., 2010).

However, most of the above studies were carried out in open, conifer or deciduous forests of relatively homogenous structures. The suitability of LiDAR data for delineating the structure of complex forest types, particularly for cool temperate rainforest and neighbouring uneven-aged mixed forests in a severely disturbed landscape has hitherto remained untested. The overall objective of this study was to use airborne LiDAR data for the classification of the cool temperate rainforest and adjacent forests in the Strzelecki Ranges, Victoria, Australia. Specific objectives included the test of quantitative description of the vertical structure of the forests, compilation of LiDAR-derived variables within stratified layers, statistical analysis of the discriminatory power of the variables, classification of forest types and accuracy assessment.

2. MATERIALS AND METHODS

2.1. Study Area

The study area (located at -38°54′ latitude and 146°30′ longitude) is in the Strzelecki Ranges, southeast Victoria, Australia. The Strzelecki Ranges are an isolated series of mountains in the southern section of the Gippsland region. Prior to European settlement the Strzelecki Ranges were occupied by the Ganai (Kurnai) and Wurundjeri indigenous tribes (Boyle and Lowe, 2004) and were densely vegetated by wet forest (also referred to as wet sclerophyll forest) and cool temperate rainforest. Wet forest is most commonly dominated by Mountain Ash (Eucalyptus regnans) (Davies et al., 2002), characterised by a tall eucalypt overstorey, a broad-leaved shrubby understorey and a moist, shaded, fern-rich ground layer that is usually dominated by tree-ferns (DSE, 2005). In eucalypt-free areas, Silver Wattle (Acacia dealbata) may be locally dominant (Davies et al., 2002). Cool temperate rainforest is defined as a closed, non-eucalypt forest, which occurs in high rainfall areas and within wet forest areas which have not been exposed to fire (Adam, 1992). Myrtle Beech (Nothofagus cunninghamii) is the dominant species of cool temperate rainforest in the study area. The understorey is characterised by tree ferns and a rich epiphytic flora including Soft Tree-fern (Dicksonia antarctica), Slender Tree-fern (Cyathea cunninghamii) and Skirted Tree-fern (Cyathea X marceccens). The ground layer is dominated by a diversity of ground ferns such as Mother Shield-fern (Polystichum proliferum), Hard Water-fern (Blechnum wattsii) and Leathery Shield-fern (Rumohra adiantiformis) (Peel, 1999; DSE, 2005). These forests have experienced widespread land clearing since European settlement. Subsequent agricultural abandonment and a significant fire history have resulted in severe landscape disturbance in the Strzelecki Ranges (Noble, 1978; Legg, 1986). Rainforest is sensitive to fire and, following fire, is often replaced by forest types comprising fire tolerant species such as some eucalypts, which rely on fire to open their protective seed pods so that their seeds can germinate (Reichl, 1966; Langkamp, 1987). For example, in the study area, there was extensive regeneration of eucalypt forest following catastrophic wildfires in 1939 and 1944. The landscape has undergone further significant changes with the establishment
of large scale plantations in the area over the last half of a century. Plantations include softwood species Radiata Pine (*Pinus radiata*) and hardwood eucalypts dominated by Mountain Ash (*Eucalyptus regnans*) and Shining Gum (*Eucalyptus nitens*) (HVP, 2010). Currently, areas bordering cool temperate rainforest in the Eastern Strzeleckis are a mosaic of different land use histories involving both natural and human disturbances, and so a very complex forest structure in the remnant patches of cool temperate rainforest and adjacent forests including wet sclerophyll and plantation forests prevails. This study focuses on an area with cool temperate rainforest distribution in the Eastern Strzeleckis, which covers an area of 1.82 km² with elevations ranging between 322 metres and 448 metres.

2.2. Data

LiDAR data were collected using an Optech ALTM Gemini LiDAR system at a flying height of 1100 m between 11 and 23 October 2009. The laser scanner was configured to record up to four returns per laser pulse. The laser pulse repetition frequency was 70 kHz, with the average point density being 4 points per square metre. The laser footprint diameter at nadir was 0.3 m. The LiDAR data used for this project were documented as 0.20 m for vertical accuracy and 0.25 m for horizontal accuracy. The LiDAR data were classified into ground and non-ground points by the vendor and were delivered in binary LAS 1.2 file format.

Ecological vegetation classes (EVCs), which describe the spatial extent of vegetation species, were provided by the HVP Plantations Pty Ltd. EVC mapping was undertaken first by the interpretation of aerial photographs and the process was designed to outline vegetation patches and any obviously related patterns. The range of aerial photograph patterns was then field checked (Davies *et al*., 2002; Boyle and Lowe, 2004). EVCs are the basic regional-scale mapping unit used for forest ecosystem assessments, biodiversity planning and conservation management in Victoria. For this study, EVC data (which include data on plantation forests) were used as reference data to establish the test plots.

2.3. Methods

Using the LiDAR ground data, a one-metre resolution DEM (digital elevation model) was generated by IDW (inverse distance weighted) method in ArcGIS 10 software for the study area. Based on the EVC reference data, a total of 70 plots (30 m × 30 m) were established. There were 14 plots for each of five forest types: cool temperate rainforest (RF), naturally regenerated Mountain Ash (MA), mixed forest (MF) consisting of overstorey Mountain Ash and understorey Myrtle Beech, Silver Wattle (SW), and hardwood plantation (PT) dominated by Shining Gum (*Eucalyptus nitens*). LiDAR data were extracted for each of the plots. The laser return height above the ground was calculated as the difference between the non-ground laser returns and the corresponding DEM value (Zhang *et al*., 2011). It was these normalised laser returns that were used for analysis of the spatial distribution of vertical forest structure for the plots dominated by different forest types.

The k-means clustering algorithm, which produces a partition of the data into the k different clusters in such a way that all individuals in a cluster are closer to their own cluster mean (Burns and Burns, 2008), was performed on each of the plots to stratify the vertical forest structure into layers, representing the overstorey (layer 1), mid-storey (layer 2) and lower storey (layer 3) of the plot-level forests. Eighteen variables were then calculated from the LiDAR data based on the three-layered structure for each plot, six variables for each layer. The variable names and descriptions are listed in Table 1.

The penetration rates were calculated by the following formula:

\[
Penetration_i = \frac{\sum_{j=1}^{4} n_j}{\sum_{j=1}^{4} n_j}
\]

where, \(i = 1, 2, 3\), representing layer 1 (overstorey), layer 2 (mid-storey) and layer 3 (lower storey); 4 represents the ground; and \(n_1, n_2, n_3, n_4\) are the number of laser returns from layer 1, layer 2, layer 3, and the ground, respectively (Zhang *et al*., 2011).

One-way ANOVA (analysis of variance), a statistical technique to test whether the observed differences between the sample means are of such magnitude as to indicate that they could have come from the same or different populations (Walford, 2011), was employed here to explore the possibility of distinguishing forest types using LiDAR-derived variables. In one-way ANOVA, the hypotheses to be tested for a variable are:

\(H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5\) (the means of a variable from the five forest types are equal)
The post hoc Tukey HSD (honestly significant differences) test (Burns and Burns, 2008) was carried out to identify which specific pairs of forest types could be discriminated for a specific variable. For each variable, the post hoc tests compared all different combinations of paired forest types to identify which forest type was significantly different from the other based on the following hypotheses:

1. The two means are equal

2. The two means are not equal.

The results from the post hoc test were used to evaluate the significance of variables for the discrimination of forest types (Suratno et al., 2009; Seltman, 2010). Linear discriminant analysis with cross-validation was performed to classify the forest types and assess the classification accuracy. The cross validation, often termed a jack-knife classification (Burns and Burns, 2008), successively classifies all forest plots but one to develop a discriminant function and then categorizes the plot that was left out. This process was repeated with each plot left out in turn (Burns and Burns, 2008).

3. RESULTS AND DISCUSSION

Figure 1 shows examples of plot-based height profiles (one for each forest type), depicting the tree heights and the vertical distribution of the laser returns at 1-m height intervals. The rainforest plot has maximum tree heights of about 30 m. The mountain ash forest has trees with maximum heights of nearly 60 m. The mixed forest shows the bimodal height distribution reflecting the taller mountain ash trees above the rainforest trees. The plantation forest shows a lower height distribution than the mountain ash forest. In terms of the variation of frequency (the number of the laser returns at different heights), the profiles of all five plots are characterised by a three-layered vertical structure. The k-means clusters stratified the vertical structure to three layers as shown in Figure 1, representing overstorey, mid-storey and lower storey.

The F-statistics and p-values for all variables from the one-way ANOVA at α = 0.05 level are shown in Table 2. All the F-values are greater than the F-critical value 2.519 and p-values are less than 0.05. Therefore, the null hypotheses for all 18 variables were rejected. The results of the one-way ANOVA test revealed that at least two of the forest types can be discriminated from each other if using any one of the 18 variables. In the post hoc tests, the variables MaxH, MeanH and Density in the overstorey, MaxH, Depth, StdDev and Penetration in the mid-storey and MeanH, Density and Penetration in the lower storey were found to be
significantly different between one forest type and at least other two types. Therefore, the combination of these 10 variables exhibited the most discriminatory power.

![Figure 1](image-url)

**Figure 1.** Examples of plot-based height profiles (one for each forest type), showing the height distribution of the laser returns within the forest plot at 1-m height interval. The k-means clusters stratified the vertical structure into three layers, with the horizontal lines in the profiles showing the boundary between the layers.

**Table 2.** $F$-statistics and $p$-values from LiDAR-derived variables from one-way ANOVA at $\alpha = 0.05$ level.

<table>
<thead>
<tr>
<th>Variables</th>
<th>MaxH</th>
<th>Depth</th>
<th>MeanH</th>
<th>StdDev</th>
<th>Density</th>
<th>Penetration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overstorey</td>
<td>$F$</td>
<td>96.879</td>
<td>26.781</td>
<td>65.861</td>
<td>28.642</td>
<td>20.190</td>
</tr>
<tr>
<td></td>
<td>$p$-value</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Mid-storey</td>
<td>$F$</td>
<td>49.833</td>
<td>67.726</td>
<td>21.330</td>
<td>52.796</td>
<td>34.124</td>
</tr>
<tr>
<td></td>
<td>$p$-value</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Lower Storey</td>
<td>$F$</td>
<td>5.769</td>
<td>5.216</td>
<td>16.885</td>
<td>4.323</td>
<td>16.434</td>
</tr>
<tr>
<td></td>
<td>$p$-value</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

**Table 3.** Classification results by the linear discriminant analysis using the ten variables as identified from the post hoc tests for different forest types.

<table>
<thead>
<tr>
<th>Forest types</th>
<th>RF</th>
<th>MF</th>
<th>SW</th>
<th>MA</th>
<th>PT</th>
<th>Row Total</th>
<th>User’s Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classified</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forest types</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>11</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>78.6%</td>
</tr>
<tr>
<td>MF</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>100%</td>
</tr>
<tr>
<td>SW</td>
<td>3</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>78.6%</td>
</tr>
<tr>
<td>MA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>14</td>
<td>100%</td>
</tr>
<tr>
<td>PT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>14</td>
<td>100%</td>
</tr>
<tr>
<td>Column Total</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>

Producer’s Accuracy = 78.6% 100% 78.6% 100% 100%

Overall Accuracy = (11+14+11+14+14)/70 = 91.4%

The results of linear discriminant analysis using the 10 variables identified as important during the post hoc tests are shown in Table 3, which is referred to as a classification error matrix (also known as a confusion table) showing the number of forest plots correctly and incorrectly assigned to each of the forest types based on the discriminant analysis (Burns and Burns 2008). The results showed that 100% of the plots dominated
by mixed forest, mountain ash and plantation were correctly classified in both the original and the cross-
validated classification results. These three forest types can be easily identified using the selected LiDAR-
derived variables. Rainforest and silver wattle forest were relatively difficult to distinguish using the linear
discriminant analysis, with up to three plots of each types being misclassified, showing 78.6% for both
producer’s accuracy and user’s accuracy. The overall results are very good, showing an overall classification
accuracy of 91.4%.

Unlike the interpretation of aerial photographs (which see forests only from above the overstorey) and field
identification using the floristic keys (which see forests from the ground), airborne LiDAR data used in this
study provide detailed information not only from the top of the canopy, but also through the interior of the
canopy including the mid-storey and the lower storey, and the ground because of the LiDAR penetration of
forest canopies (Zhang et al., 2011). This study demonstrated the applicability of airborne LiDAR for the
classification of Australian cool temperate rainforest and adjacent forests. LiDAR technology has advantages
over traditional methods in forest type classification. Statistical analyses showed that the LiDAR-derived
variables have significant discriminatory power to distinguish between forest types. Given that the historical
and ecological significance of cool temperate rainforest in Victoria, Australia is such that endangered EVC
status has been bestowed, these results are of practical importance. Rainforests in Victoria are protected from
the impacts of timber logging through the use of appropriate buffers to maintain microclimatic conditions and
protect from disease and other disturbance. It is here that boundaries between rainforest and adjacent forests
must be well positioned and monitored so that buffer zones requisite for effective rainforest conservation and
protection can be agreed upon (Cameron, 2008). In addition, vertical description of the forest structure is
essential to study biological diversity, microclimate, ecological processes, succession, and the history of
natural and anthropogenic disturbances of the landscape. Therefore, there will be great potential for
applications of LiDAR data in sustainable forest management and native forest conservation.

4. CONCLUSION

This study used airborne LiDAR data to describe the vertical structure of forests in one of the cool temperate
rainforest areas in Victoria, Australia and tested the capability of LiDAR-derived variables for distinguishing
and classifying cool temperate rainforest and neighbouring forest types. Using laser point data normalized by
ground elevation in this study, the forest structure was stratified into layers representing the overstorey, mid-
storey and lower storey of the forests within each plot. The variables derived from the LiDAR data were
based on each of these layers rather than on each level with a predefined height interval. These variables are
found to be effective in characterising the vertical structure of the forests examined. Statistical analyses
including one-way ANOVA and the post hoc tests were used to evaluate the significance of variables for the
discrimination of forest types. Using linear discriminant analysis with the cross-validation, an overall
classification accuracy of 91.4% was achieved in the study area.

REFERENCES

bioregion, pp.142. Victorian Department of Sustainability and Environment, Melbourne, Australia.
Heidelberg.
delineation of rainforest boundaries, pp.139. Victorian Department of Sustainability and Environment,
Melbourne, Australia.
1:25000 in Gippsland, Victorian Department of Natural Resources and Environment, Melbourne,
Australia.
DSE (2005). EVC Benchmarks - Strzelecki Ranges bioregion, Victorian Department of Sustainability and
Environment, Melbourne, Australia.
Forestry Chronicle, 82(2), 211-218.
Zhang et al., Statistical analysis of airborne LiDAR data for forest classification in the Strzelecki Ranges


Peel, B. (1999). Rainforests and Cool Temperate Mixed Forests of Victoria, Department of Natural Resources and Environment, East Melbourne, Victoria, Australia.


