HERE, THERE, OR EVERYWHERE? AN ACCURACY ASSESSMENT AND THE INFLUENCE OF ECOREGIONS IN THE SPATIAL DISTRIBUTION OF ERRORS IN A VEGETATION LAND COVER MAPPING ASSESSMENT OF VICTORIA, AUSTRALIA.

Peter King[1], Simon Jones[1], Mariela Soto-Berelov[1], Andrew Mellor[2] and Andrew Haywood[2]

[1]Mathematical and Geospatial Science RMIT University, Melbourne, Victoria, Australia

[2]Department of Sustainability and Environment, Victoria, Australia

Corresponding author: Peter King, s3344741@student.rmit.edu.au

Forests provide many benefits to individuals, societies as well as the biosphere in general. Forest loss can lead to increased greenhouse gases, an increase in salinity, decreased biodiversity and the extinction of species. As part of the sustainable management of Victoria's forests the Department of Sustainability and the Environment (DSE) has created the 2008 land cover baseline via the aerial photographic interpretation (API) of 790 2x2 km digital high resolution colour aerial photographs, so that information regarding the state, extent and condition of Victoria's public land forests can be monitored. This study conducted an accuracy assessment of the resultant API maps, with respect to their ability to classify forested/ non-forested land cover, forest height, and forest vegetation type. This assessment was made by comparing the API derived maps to 125 field plots. An aspect of the spatial distribution of API errors was investigated by stratifying the data set into three Victorian ecoregions. Results indicate that the API accurately classified forest/ non-forest land cover and vegetation type. The classification of height was less precise, with the Mediterranean forest, woodland and scrub ecoregion obtaining more successful results. Despite these errors, the API was successful at identifying forest areas and is a valuable tool in the monitoring and reporting of forests across Victoria.

KEYWORDS; API, accuracy assessment, ecoregion, vegetation, Victoria.

Peter King is a geospatial science postgraduate student at RMIT with an undergraduate BSc in ecology and biodiversity. He has published in the journal Population Ecology with research investigating the possibility of seed dispersal by Wellington Tree Weta.

Simon Jones is the Director of the Remote Sensing and Photogrammetry Research Centre at RMIT. His areas of expertise are remote sensing, ground verification, spatial analysis, spatial data uncertainty, land-cover mapping, monitoring & modelling, and vegetation.

Mariela Soto-Berelov is a post doctoral research fellow in remote sensing at RMIT (SMGS). She specializes in environmental geography, land-use change science (LUCC), vegetation mapping/ modelling, geographic information science, and remote sensing.

Andrew Mellor & Andrew Haywood are with the forest monitoring and reporting section of the Department of Sustainability and Environment, Victoria, Australia.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

1 Introduction

Throughout history, humans have dramatically altered and transformed the landscape (Lillesand *et al.* 2004, Steffen *et al.* 2004). Changing land cover can impact the environment by altering the structure and function of ecosystem properties such as water chemistry, the carbon cycle, the water and energy exchange balance, and atmospheric composition and chemistry (Steffen *et al.* 2004). Land cover change is considered at least as significant as climate change (Foody 2002) and is predicted to be the most important factor affecting biodiversity for at least the next 100 years (Chapin et al. 2000).

Forests provide a host of benefits and the loss of forests can result in increased greenhouse gases, increased erosion, changed salinity, decreased biodiversity, and cause the extinction of species (Lucas *et al.* 2006, Steffen *et al.* 2004). In order to protect and manage the temperate and boreal forests of the world in a sustainable way, 12 countries, including Australia, formed a working group titled the Montréal Process in 1994. Using the guidelines produced by the 1992 United Nations Conference on Environment and Development (UNCED) the Montréal Process developed criteria and indicators which provide advice on how to assess, describe, and evaluate progress related to the sustainable management of forests (DSE 2007, Siry et al. 2005).

To guide the management of forests in Victoria, the state government adopted seven key criteria proposed by the UNCED and the Montréal Process (see Table 1) and used these as a base to create 45 indicators of their own. This formed a platform for the measurement and reporting of the sustainable management of Victoria's forests administered through the Department of Sustainability and Environment or DSE (DSE 2007). Through the reporting process, DSE produces a five-yearly *State of the Forests Report* which provides information on the environmental, economic and social values of the forests. The *State of the Forests Report* facilitates the continual improvement of forest management by enabling the evaluation of the management performance, and this evaluation in turn can guide the enhancement of forest policy (DSE 2009).

Table 1: The seven criteria adopted by the Victorian Government (DSE 2009).

Criteria 1:	Conservation of biological diversity
Criteria 2:	Maintenance of productive capacity of forest ecosystems
Criteria 3:	Maintenance of ecosystem health and vitality
Criteria 4:	Conservation and maintenance of soil and water resources
Criteria 5:	Maintenance of forest contribution to global carbon cycles
Criteria 6:	Maintenance and enhancement of long term multiple socio-economic benefits to meet the needs of societies
Criteria 7:	Legal, institutional and economic framework for forest conservation and sustainable management

The Victorian Forest Monitoring Program (VFMP) established by DSE has three main components: state-wide mapping using moderate-resolution remote sensing imagery, high resolution mapping via aerial photographic interpretation (API), and a systematic ground based assessment using a stratified grid of forest monitoring plots. The second high spatial resolution mapping component (the validation of which is the focus of this paper) involved the creation of a 2008 land cover baseline (2008 LCB) using API of non-stereo, digital aerial photographs. The 2008 LCB data will provide forest classification and land cover area estimates for the VFMP. It will also contribute to the monitoring and reporting of the state, extent and condition of Victoria's public land forests. The 2008 LCB was required to map broad forest types and structural components (based on the National Vegetation Information System or NVIS), as well as non-forest land cover types (based on the UN FAO Land Cover Classification System or LCCS). During this process, the following attributes were delineated and classified: dominant forest or land cover class, canopy height and canopy cover.

The use of remote sensing to monitor forests is not unique to Victoria as remote sensing techniques offer an ability to map forests synoptically across large areas as well to characterise forests attributes (e.g., biophysical, structural, biochemical, disturbance) over a range of scales and time intervals at a relatively low cost (Linke *et al.* 2006). Additionally, benefits of using aerial photography include the following: it can be gathered at any time or place; is often of high spatial resolution; involves less atmospheric interference; and can provide long time series from as far back as the 1930's (Morgan, Gergel & Coops 2010). As such, remote sensing (aerial photography in particular) is invaluable for monitoring forests and detecting land cover change and can ultimately assist the sustainable management of forests.

1.1 Accuracy Assessment

Accurate maps of land cover and its elements are required to assist in the sustainable management of forests. Given all maps contain a level of generalization of the real world and thus a loss of information, they inherently contain a certain degree of error. It is therefore necessary to provide an assessment of the accuracy with which land cover and its elements are mapped so that not only the quality and appropriateness of the map is understood, but also any error and the potential consequences of this error (Foody 2002).

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

Accuracy assessment is considered a fundamental part of land cover mapping. However, there is no standard method for the reporting of accuracy. This is partly due to the varied nature of land cover maps, as well as the variety of problems associated with reporting accuracy assessments, such as what is the necessary information and how to convey it (Congalton & Green 2009, Foody 2002). Foody (2002) provides a detailed summary of the problems associated with the accuracy assessment of land cover classifications. The author concludes that a sound approach to the reporting of classification accuracy should include a confusion matrix and at least two quantitative measurements such as overall accuracy and the kappa statistic, along with confidence levels (Foody 2002). The recommendations of Foody (2002) are followed in this report.

Whilst it is important to report thematic accuracy (or the ability of a map to classify different categories), this fails to take into account variation in the spatial distribution of error which often occurs (Foody 2005, McGwire & Fisher 2001). For instance, a study by Foody (2005) reported an overall accuracy of 84%, even though local accuracy ranged from 53 to 100%. Spatial variation in misclassifications have been known to be produced by sensor properties (Foody 1988), and they also frequently occur at class boundaries due to the presence of mixels (a pixel containing more than one class of object) and misregistration (Congalton 1988, Steele *et al*. 1998, Vieira & Mather 2000). As a consequence of these studies, it can be concluded that raw confusion matrices fail to represent the spatial variability of classification error. To address this, different approaches have been developed. The majority of these focus on mapping a measure of the uncertainty involved in labelling a class on a per pixel basis. Such approaches allow the operator to pinpoint where on a map there are high levels of uncertainty and guide investigations into why this spatial variability in uncertainty may occur (Stehman & Foody 2009).

Another more general approach which has been undertaken involves dividing the map into subregions and performing individual accuracy assessments on each of these subregions via standard confusion matrices and estimates of overall accuracy (e.g. Foody 2005). By investigating and reporting on the variability of errors between spatial distributed sub-regions, researches are able to locate problematic areas and thus refine the classification methods to increase their accuracy (Foody 2002). Another benefit of this approach is that it can usually be undertaken using the data which has already been collected, thus eliminating the need to acquire more data which is typically costly (Foody 2005).

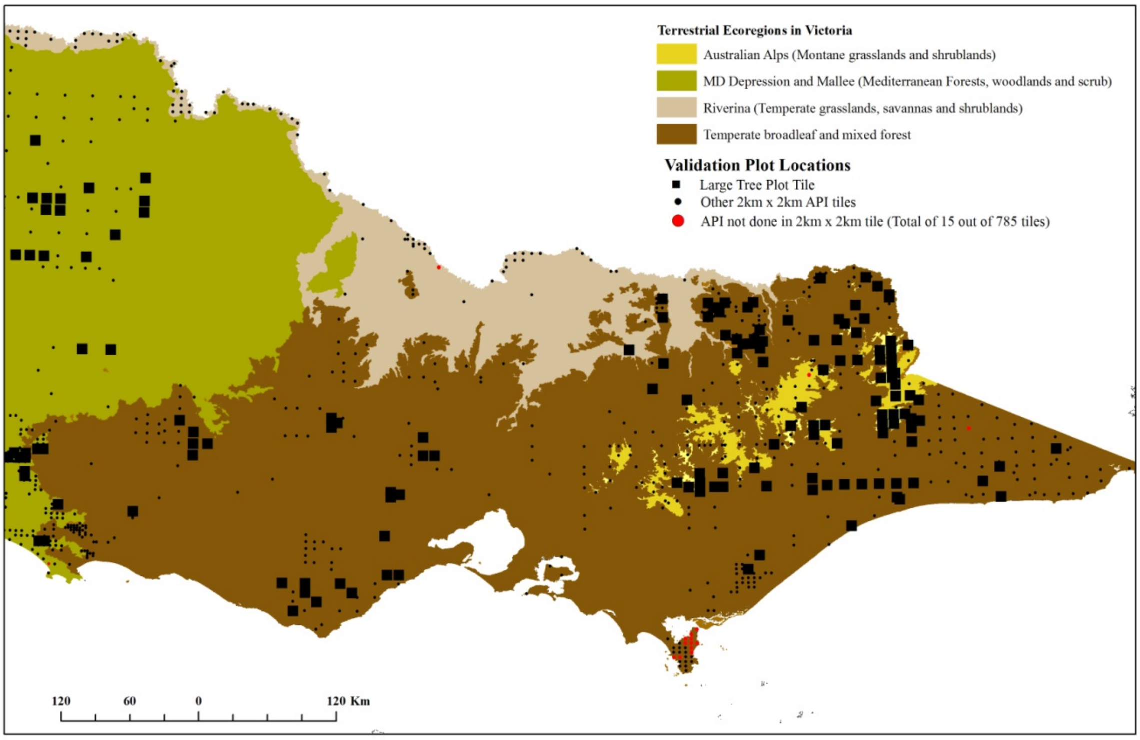

Figure 1. Location of API tiles across Victoria as well as those used in this study for validating the API using the LTP survey data. All large tree plots also contain an API photoplot and are plots used in this study.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

2 Methods

The aim of this study is to validate the API used to create the 2008 LCB, with the goal of using this knowledge to improve the mapping process in the future, if needed. It addressed the API's ability to classify forest versus non-forest land cover, vegetation type, and height. It also investigated an aspect of the spatial distribution of errors according to ecoregions, and examined potential sources for these.

Ecoregions are a "relatively large unit of land or water containing a characteristic set of natural communities that share a large majority of their species, dynamics, and environmental conditions" (Olson & Dinerstein 1998). The ecoregion approach in grouping environments is used by the World Wildlife Fund (WWF) to improve conservation by ensuring that all ecosystem and habitat types are described in their Global 200 and thus represented by conservation strategies throughout the world (Olson & Dinerstein 1998). These ecoregions are also important in Australian conservation as they are the foundation of the 89 geographically distinct bioregions identified in the Interim Biogeographic Realisation for Australia or IBRA, and these bioregions are a key tool in the identification of land for conservation in Australia's Strategy for the National Reserves System 2009- 2030. Whilst it would have been preferable to use the IBRA bioregion subdivision in this study (as these were used to stratify the sampling grid for selecting photoplots), there was insufficient data points available for this approach. The current analysis aimed to determine whether the different ecoregions present in the study area had any influence on the accuracy of the API. This was achieved by dividing the study area into ecoregions and assessing these ecoregions individually.

2.1 Creation of the API

The API mapped state forests and parks across Victoria during 2008, an area encompassing approximately 7.8 million hectares of forest (DSE 2009). This involved the interpretation of 790 2x2km digital (georectified) high-resolution (≤ 50 cm) colour aerial photographs or photoplots, taken between 2006 and 2010, (see Figure 1). The photoplots were selected from a state-wide sampling grid which had been stratified according to IBRA bioregions, with an average of 72 photoplots in each bioregion. Classification of the photoplots involved two main stages; the delineation of landscape elements and their subsequent manual verification and classification, with a minimum mapping unit (MMU) of 0.5 ha.

The delineation of landscape elements conducted during the first stage relied on an automated segmentation algorithm applied using *Definiens eCognition* (version 3). This algorithm employs a region-growing technique where pixels are amalgamated into larger objects, only stopping if the smallest growth exceeds the heterogeneity threshold. This threshold is defined by segmentation parameters which combine two criteria. The first relates to the homogeneity of the aerial photograph pixel values and the second to the shape (compactness and smoothness) of the resultant image object.

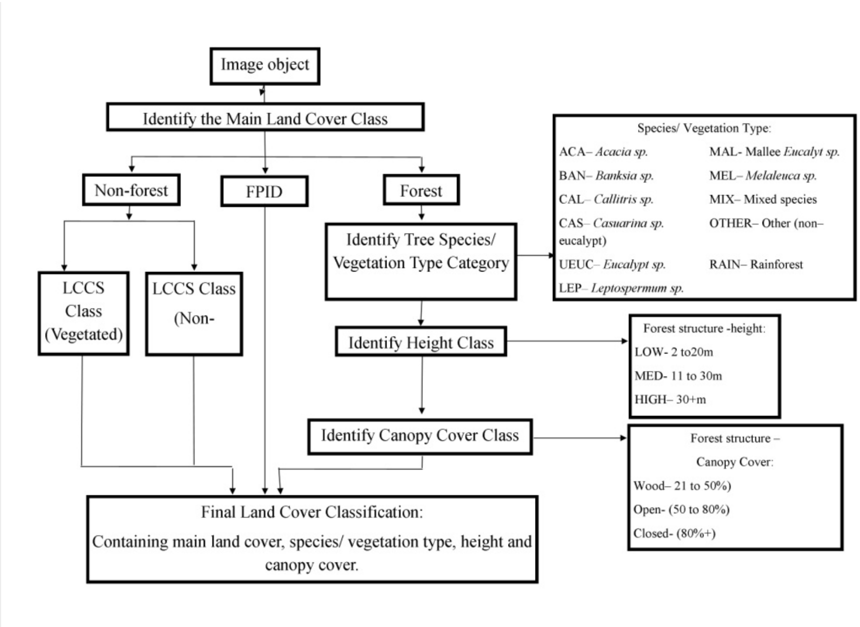

Figure 2. An outline of the forested land cover classification used for the API (adapted from Farmer *et al.*2011). If at any stage the majority of the object could not be placed into a single category, it was split (so long as the MMU is greater than 0.5 ha). If an area had substantial fire damage it was classified as FPID (Fire Prevents Identification) and no further characterisation was untaken.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5

The second stage, the manual verification and classification of the landscape elements, utilised tone, colour, shadow, size, shape, texture, and overall context of the photoplots. Ancillary data including fire histories and modelled vegetation types were also used for reference. All interpreters were provided with a series of interpretation manuals, field visits, and expert consultants to assist them during the initial stages of the project. A hierarchical land cover classification approach was applied, with the top tier comprising of broad land cover types. For a landscape element to be classified as forest it was required to be larger than 0.5ha with a canopy height higher than 2m, and greater than 20% canopy cover. All land cover identified as forest was subsequently assigned vegetation type, height and canopy cover (see Figure 2 for an outline of the classification and the classes involved). Vegetation type was defined as the dominant top canopy species vegetation type (see Figure 2), and stands containing multiple, co-dominant broad species types were classified as non- *Eucalypt* or mixed.

2.2 Reference data used to assess the accuracy of the API

The accuracy assessment of the API conducted in this study relied on the third tier of data used in the VFMP: the systematic ground based assessment which represents the highest level of detail. The third-tier of the FM & RIS contains approximately 142 installed and measured large tree plots or LTP. These plots were measured using standard operating procedure 13 (DSE 2011). The LTP are located at the centroid of the 2x2km photoplots, and comprise a circular plot of 0.04 ha. For each LTP, a range of metadata was recorded including individual tree attributes such as tree species, diameter at breast height (DBH) for all trees larger than 10cm, heights across a range of DBHs, and crown class. Forest stand attributes such as canopy cover and administrative essentials such as the person(s) involved and the date, were also recorded. The dominant top canopy was used when assigning vegetation type.

In the current study, the LTP data was used to assess the classification of the API. Specifically, the forest/non-forest land cover, vegetation type (main vegetation type mapped), and height class results obtained from the API were compared against the LTP to determine the API's accuracy. The different classes of non-forest land cover were not investigated as the API has primarily been produced to record forest cover as opposed to land cover in general, and the low number of non-forest plots made any analysis of the classification of non-forest land cover difficult.

Nine of the 142 LTPs were not utilized. Circumstances which hindered the potential to compare these nine LTPs against their associated API classification include: a disturbance event such as logging had recently occurred; the photoplot was classified during the API as fire prevents identification or FPID (8 tiles), a term used to identify areas where fire had prevented the confident attribution of species and/or height and/or canopy cover in the original identification; and missing or incomplete data in the LTP (2 tiles). Consequently 125 LTPs were available for comparison against the API for the forest/non-forest land cover analysis and the vegetation type analysis. This was reduced to 123 plots in the height analysis.

2.3 Analysis of the API versus reference data

This study assessed the accuracy of the API in classifying forest/ non-forest land cover, vegetation type, and height class by comparing the classifications of the API for each landscape element with the findings of the corresponding LTP which was located at the centroid of the landscape element. This was conducted for 125 photoplots. This report was created to be presented at the GSR 2 conference (Dec 10-12, 2012), as such the length and scope was limited which resulted in canopy cover not being analysed. As there are known issues in classifying canopy cover, via API (Farmer *et al.* 2012), it is assumed this class may have obtained similar results to those of the interpretation of forest stand height (see subsequent height class analysis).

2.4 Accuracy assessment

Confusion matrices were produced for forest/ non-forest land cover, vegetation type, and height class along with total accuracy, total accuracy using the true marginal proportions to correct for bias (along with confidence intervals), the kappa statistic, producer and user accuracy. Unfortunately, low frequencies of samples in certain categories meant this was not always possible. The matrices were not normalized as this can significantly bias the accuracy assessment (Stehman 2004). The benefit of the total accuracy measure which utilises the true marginal proportions, that is, the proportion of the map which falls into each category, is that it corrects for bias when calculating total accuracy statistics. Consequently, the size of the polygons (in this case the landscape elements) and thus bias caused by irregular sized polygons is removed. Total accuracy, as opposed to the total accuracy using true marginal proportions, has been compared in the discussion since out of the two statistics, it is more common and readily understood.

2.5 Assessment of an aspect of the spatial distribution of error

This study has investigated one aspect of the spatial distribution of error, namely the variability in errors between ecoregions. This is in no way a comprehensive assessment of the spatial distribution of errors, however it is the belief of the authors that it may aid in locating problematic areas and thus help refine the classification methods. There are four WWF global ecoregions, across the state of Victoria. Three of these were used in this study to investigate if accuracy was equally distributed throughout the study area. The three ecoregions used are: Ecoregion 1, Mediterranean forest, woodland and scrub; Ecoregion 2, Montane grassland and shrubland; and Ecoregion 3, Temperate broadleaf and mixed forest,(see Table 2 for a description of these ecoregions and Figure 1). The fourth ecoregion in the study area, the temperate grasslands, savannas, and shrublands, was not used because

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

unfortunately the LTP dataset did not contain observations in this ecoregion. In the analysis of the spatial distribution of error variation has been attributed to differences arising from qualities (such as vegetation characteristics) of the ecoregions whilst it should be noted that elements (such as image quality) could also be responsible.

Table 2. Descriptions of the three WWF global ecoregions involved in this study. (www.environment.gov.au/parks/nrs/science/bioregion-framework/terrestrial-habitats.html)

Ecoregion	Description	Designation
Mediterranean forest, woodland and scrub	Defined by cool and moist winters, and summers which are hot and dry. This ecoregion contains many unique plants and animals, and numerous plants in particular are fire adapted relying on the disturbance the fires bring.	Ecoregion 1
Montane grassland and shrubland	Encompasses grassland and shrublands at high elevations, and covers less than 3% of Australia's land area.	Ecoregion 2
Temperate broadleaf and mixed forests	A region which is variable in the temperature and amount of rainfall. Composed predominantly of eucalypt and acacia species.	Ecoregion 3

2.6 *Fuzzy accuracy assessment of the height class*

Unlike vegetation type and forest/non-forest, the height class is not a nominal category and is instead a continuum. This ambiguity can result in error being perceived in the analysis due the assigning of divisions, when in fact the misclassified plots lie close to the boundaries and could be assigned to either class. For instance, if a tree which was measured in the LTP as being 30m high was placed in the high height category it would be perceived as being wrongly assigned when in reality it lies so close to the boundary of the categories that it could be considered to fall within the medium or high classes. Fuzzy analysis techniques try to mitigate this by allowing objects and pixels to belong to more than one class (Lillesands *et al.* 2004, Lunetta & Lyon 2009). Because of this, fuzzy logic was introduced and two additional classes were added to the analysis (low-medium and medium-high), creating transition zones where a point was deemed to be accurately classified if it fell within the LTP class or the class adjacent to this category. For instance, if the reference data plot was in the medium class then it could fall into the medium, low-medium, or medium high class. On the other hand, a low reference plot could be assigned to the low or low-medium class and be deemed accurate (see Table 3 for the ranges used in the two fuzzy height analyses). Continuing with the previous example, a 30m tall tree would now belong in the med-high class and would be considered to have been accurately classified if it was assigned to either the medium or the high class in the API.

Table 3. Height classes used in the fuzzy height analyses.

	Height class				
	low	low-medium	medium	medium-high	high
Original height analysis	2-10m		11-30m		31m+
Fuzzy accuracy analysis 1 (a 1m transition zone)	2-9m	10-11m	12-29m	30-31m	32m+
Fuzzy accuracy analysis 1 (a 4m transition zone)	2-6m	7-14m	15-26m	27-34m	35m+

3 Results

Due to the nature of the data not all categories contained sufficient samples for accurate analysis, and thus the producers and users accuracy has not been reported for a number of categories. This is particularly evident during the later analyses as the classes are further sub-divided producing lower sample populations. The kappa statistic has been reported in order to maintain transparency, even though a number of these are not appropriate representations of the accuracy. Kappa is a ratio (it is dependant on the numerator and denominator), and classes with an uneven distribution can be problematic during the calculation and interpretation of kappa (Pontius & Millones 2011). This study frequently involved asymmetrical classes, such as the occurrence of more forested areas than non-forest (as the mapping mostly involved state forests (Farmer *et al.* 2011)), and these have resulted in a number of kappa statistics which are indefinable or zero, when the classification was highly accurate, such as the analysis of the prediction of main land cover for ecoregions 2 and 3.

3.1 *Forest/ non-forest*

Results that report on the accuracy of the Forest/Non-Forest classification are shown in Tables 4 and 5 and Figure 3. The API was highly accurate in classifying woody areas with a total accuracy of 96.7%, and a producers and users accuracy of 99.1 and 97.3 respectively for forest/ non-forest land cover. The accuracy obtained throughout the ecoregions varied. Ecoregion 1 showed slightly more error in the classification of forest/non-forest cover than the other ecoregions obtaining a total accuracy of 88%.

Table 4. A confusion matrix for forest/ non-forest land cover comparing the classifications produced by the API (map data) and the LTP (reference data).

		Reference Data		
		Forest	Non-Forest	Totals
Map Data	Forest	113	1	114
	Non-Forest	3	8	11
	Totals	116	9	125

Table 5. Quantitative measurements produced for the accuracy analysis of the forest/ non-forest land cover classification for the entire region and individual ecoregions (after stratifying the data).

	Region			
	Overall	Ecoregion 1	Ecoregion 2	Ecoregion 3
Total accuracy:	96.8	88	100	98.55
Total accuracy, using true marginal proportions:	95.6	89.2		
95% Confidence interval using true marginal proportions	92.2-98.9	77-100		
kappa:	0.78	0.74	0	0
User accuracy Forest:	99.12	93.33	100	100
Producers accuracy Forest:	97.41	87.5	100	98.55

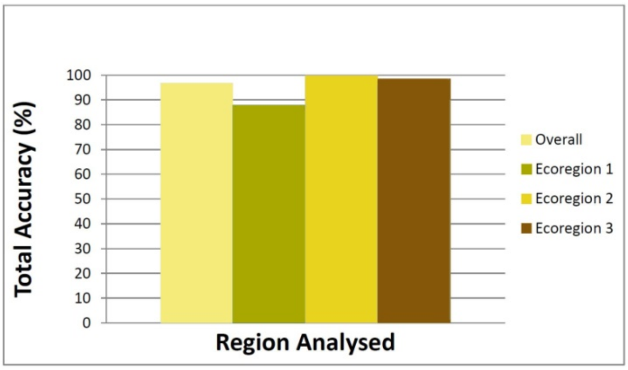

Figure 3. A histogram of the total accuracy of the forest/ non-forest land cover classification, for the overall region, and for the 3 ecoregions after stratifying the data set..

3.2 *Vegetation type*

The API was accurate in predicting vegetation type with a total overall accuracy of 88.8%, and a users and producers accuracy of 94.4 and 93.5%, respectively for *Eucalyptus* species (Tables 6 and 7 and Figure 4). There was spatial variation in error with Ecoregion 1 being the least accurate, although still achieving a reasonable total accuracy of 80%. The total accuracy using true marginal proportions obtained lower result of only 55.3%.

Table 6. A confusion matrix for the vegetation type class identified API (map data) and the LTP (reference data).

		Reference Data					
		EUC	MIX	MAL	Acacia	Non-forest	Totals
Map Data	EUC	102	4		2		108
	MIX	3	1				4
	MAL	1				1	2
	Acacia						0
	Non-forest	3				8	11
	Totals	109	5	0	2	9	125

Table 7. Quantitative measurements produced for the accuracy analysis of the vegetation type classification for the entire region and individual ecoregions.

	Region			
	Overall	Ecoregion 1	Ecoregion 2	Ecoregion 3
Total accuracy:	88.8	80	87.1	92.75
Total accuracy, using true marginal proportions:	87.3	55.3	90	94.7
95% Confidence interval, using true marginal proportions:	81.9 - 92.7	40.6-70	79.2-100	89.6-99.9
kappa:	0.53	0.63	-0.04	0.03
User accuracy EUC:	99.12	93.33	100	100
Producers accuracy EUC:	97.41	87.5	100	98.55

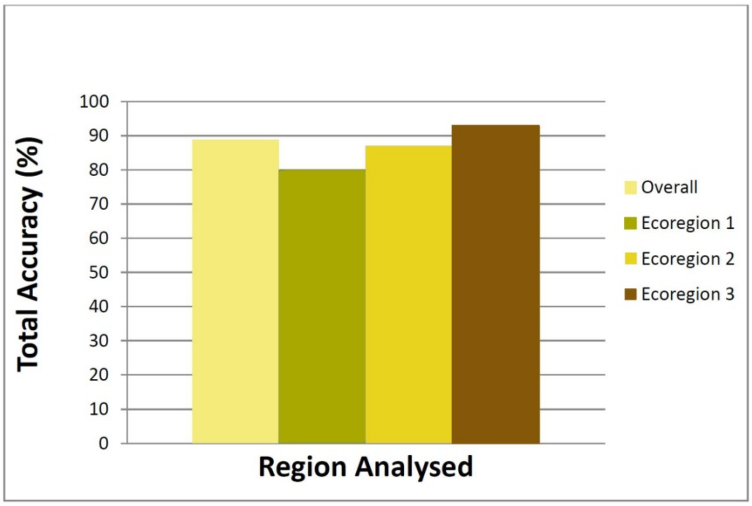

Figure 4. A histogram of the total accuracy of the vegetation type classification, for the overall region, and for the three ecoregions after stratifying the data set.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

3.3 Height

The API was not as successful in predicting the height of trees as it was in predicting forest/ non-forest (Tables 8 and 9 and Figure 5). The total overall accuracy obtained was 52%. Spatial variation in error was also observed. Ecoregion 1 obtained a total accuracy of 75%, whilst ecoregions 2 and 3 obtained 48.4% and 45.6%, respectively. The inclusion of transition zones in the fuzzy accuracy assessment increased the accuracy to 59% with inclusion of a 1m transition zone (fuzzy height 1), and 71% with inclusion of a 4m transition zone (fuzzy height 2). Both of these are still below the 85% total accuracy which is recommended in the literature (Thomlinson *et al.* 1999), although the total accuracy of fuzzy height 2 rose above the 70% minimum threshold identified by Thomlinson *et al.* (1999).

Table 8. A confusion matrix of the height class comparing classification produced by the API (map data) with the LTP (reference data).

		Reference Data				
		LOW	MEDIUM	HIGH	Non-forest	Totals
Map Data	LOW	5	6		1	12
	MEDIUM	16	47			63
	HIGH		33	4		37
	Non-forest	3			8	11
	Totals	24	86	4	9	123

Table 9. Quantitative measurements produced for the accuracy analysis of the height class classification for the entire region, the individual ecoregions, and for the two fuzzy classes.

	Region					
	Overall	Ecoregion 1	Ecoregion 2	Ecoregion 3	Fuzzy Height 1	Fuzzy Height 2
Total accuracy:	52	75	48	46	59	71
Total accuracy, using true marginal proportions:	58.3	73.6	43.5	46.7	64.5	76.9
95% Confidence interval using true marginal proportions:	50.8 - 65.8	55.6 - 91.4	28.9 - 58.2	37.8 - 55.6	57.5-71.6	72.3-81.5
kappa:	0.2	0.62	0.13	-0.03	0.39	0.49
User accuracy LOW:	41				58	83
User accuracy MEDIUM:	74		85	72	83	100
Producers accuracy LOW:	21				33	77
Producers accuracy MEDIUM:	55		44	53	59	66

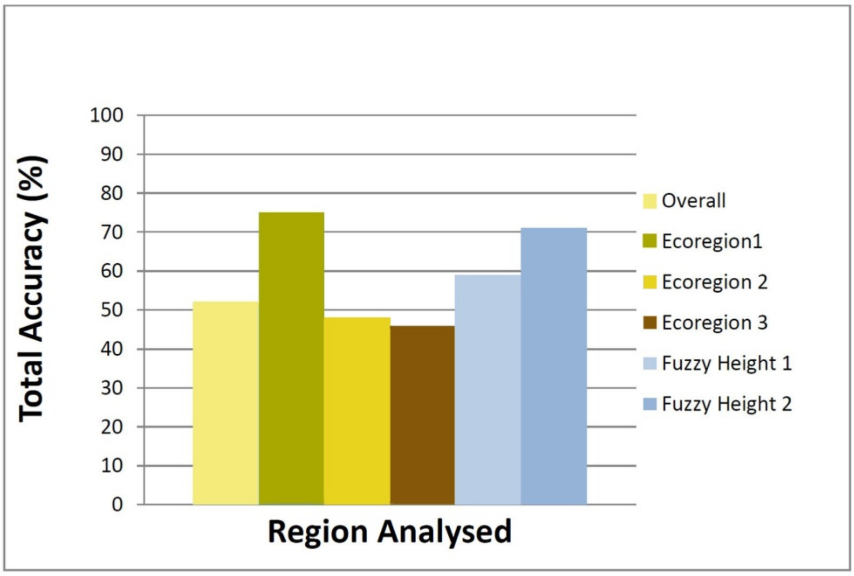

Figure 5: A histogram of the total accuracy of the height class classification, for the overall region, for the three ecoregions after stratifying the data set, and for the two fuzzy height analyses.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

4 DISCUSSION

4.1 Limitations

A limitation of this study is the use of a single point of reference data (the LTP) as a representation of an entire polygon which can potentially cover a very large area. If this point is not representative, for instance if it contained a stand of low trees in a polygon which was dominated by medium trees, this would invalidate any results utilising this point. This has been reduced by selecting the LTP from the middle of the objects instead of the boundaries where there is likely to be a combination of classes.

4.2 Forest/ non-forest classification

The API was highly accurate in classifying forested areas as demonstrated by the accuracy assessment conducted in this study. It is important to note that the goal of this classification was to describe state forests and parks across Victoria therefore there is an inevitable bias, in the LTPs and API photo-plots, towards forested areas (generally speaking, few landscape elements were located in non-forested areas as all LTPs fell inside of state parks and forests). Nevertheless, the API was successful at classifying forested areas. Such high accuracy is favourable given further classifications such as vegetation type and height depend on the correct classification of forest/non-forest areas.

A possible explanation for the misclassified landscape elements (four in total) is a lack of visual clues. As previously discussed, visual clues, such as tone, colour, shadow, size, shape and texture, were used when manually interpreting the aerial photography and a lack of these would have hindered the accurate classification. In particular low trees which lacked shadows, and trees with smaller crowns may have hindered forest identification and classification. All the misclassified landscape elements which were misclassified were either within the low height class and misclassified as non-forest (3 of the 4 misclassifications) or (2) misclassified non-forest as forest. Furthermore, in one of these instances, the LTP indicated the presence of dead crowns which may have caused some difficulty when interpreting the aerial photography. These errors may have been avoided had there been more visual clues, higher resolution or supplementary images, or supplementary LiDAR data which provides canopy height information.

The accuracy obtained throughout the ecoregions varied. Ecoregion 1 showed slightly more error in the classification of forest/non-forest cover than the other ecoregions. The average tree height in Ecoregion 1 was 6.1m, whilst the average height in ecoregions 2 and 3 were 20.6m and 14.7m, respectively. The lower accuracy obtained in areas where trees have low heights suggests that the height of forest stands may impact the classification of forest/non-forest areas, with areas of low tree heights being harder to classify. The use of additional ancillary data and/or supplementary imagery may lessen this issue.

In summary, the API was found to be successful at classifying forest/ non-forest land cover. A spatial variation in error was detected with ecoregion 1 obtaining the lowest accuracy of 88%, although still within reasonable limits (Thomlinson *et al.* 1999).

4.3 Vegetation type classification

The API was accurate in classifying vegetation type with a total overall accuracy of 88.8%, and a users and producers accuracy of 94.4 and 93.5%, respectively for *Eucalyptus* species. Unfortunately, other vegetation categories were not represented in a statistically significant way within the LTPs. Consequently this study was unable to assess the accuracy achieved when classifying other vegetation categories, for example *Acacia sp., Banksia sp., Casuarina sp., Melaleuca sp.* via API. The high accuracy achieved with the *Eucalyptus sp*. category was expected as Eucalyptus accounts for 93% of all trees throughout the Victorian state forests (DSE 2009). Of the classes considered in the API, the second most common is the *Mallee Eucalyptus* species which makes up 19% of the Victoria's state forests. The third most common vegetation type found in Victorian state forests is the *Casuarina* species at 2%. Given the high rate of *Eucalyptus*, it is unlikely the additional 358 LTP's that are yet to be processed will contain sufficient non-Eucalypt sp. points to test the API's ability to classify vegetation type beyond *Mallee sp.* and *Eucalyptus sp*.

Authors argue that not all errors are equal as some may be more important or damaging. Equally, it is also common to make mistakes when classifying similar classes (Foody 2002). In this study, the majority of errors (8 out of 11) which occurred when classifying vegetation type involved incorrectly assigning the vegetation type as *Eucalyptus sp.* when it was a mixed stand of *Eucalypt* and other species, or vice versa. In one instance, a plot misclassified as *Eucalyptus sp.* should have been assigned to the category *Mallee Eucalyptus sp.*, which is still a geographical sub-division of the *Eucalypt* species. The misclassification of stands of *Eucalyptus*, *Mallee Eucalyptus*, and *mixed stands* (that is, stands containing *Eucalyptus* and at least one other species) is not as detrimental as other types of misclassifications, since they are mapping types of forest which may provide similar ecosystem services such as the level of carbon storage. It is not surprising that such misclassifications are the most common errors due to the similarity of these vegetation types.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

Spatial variation in the classification of vegetation type was evident with Ecoregion 1 being the least accurate, although still achieving a reasonable total accuracy of 80%. In Ecoregion 1, two out of fourteen points (14%) correctly identified as forested vegetation were misclassified. Furthermore, these two points represented the only observations in the analysis where a vegetation type other than *Eucalyptus* was, or should have, been classified. Whilst the miss-classification of non-*Eucalypt* species is likely to also occur in other ecoregions, and some may argue that a better solution for improving non-eucalypt classification would be to focus on technique which improve this classification of non-*Eucalypt* species, investigations into why this region achieved the lowest accuracy (although still reasonable at 80%), and contained the most significant errors where non-eucalypt vegetation types were misclassified, may be beneficial. The total accuracy using true marginal proportions was significantly lower than the total accuracy in Ecoregion one suggesting the plots which have been misclassified are quite large and require a review. These plots may also benefit from LTPs to determine if the LTPs were representative of the forest.

In summary, the API was very accurate at classifying vegetation type within forests with the least successful ecoregion still obtaining 80% total accuracy. The classification of the *Eucalyptus* vegetation type was very successful obtaining a users and producers accuracy of 94.4 and 93.5% respectively. However, due to a lack of observations that are non-*Eucalypt*, this study was unable to test the API's ability to classify non-*Eucalyptus* species. Nevertheless, the majority of errors were not of a high magnitude since they involved very similar vegetation types, that is, within the *Eucalyptus* species classes (*Eucalyptus sp. Mallee Eucalyptus sp*., and mixed stands). Confusion of the Eucalypt and non-Eucalypt vegetation type was evident. Therefore, investigations that look into improving the API's ability to classify vegetation type would benefit by focusing on means to improve the identification of non-*Eucalypt* species via API or assessing why there were differences detected with regard to accuracy obtained across the different ecoregions.

4.4 Height classification

The API was not as successful in classifying the height of trees as it was in classifying forest/ non-forest and vegetation type. Two aspects potentially explain this and could help guide improvement in this area. Aspects which had the potential to inhibit the API's ability to classify height are: (a) the ambiguous nature of height classification when using aerial photography, (b) the use of non-stereo, on-screen photogrammetry to classify height.

The fuzzy analysis, which was run in order to take into account the ambiguous nature of height classification, found that a 1m transition zone did not significantly increase the classification accuracy (see Table 9). Furthermore, a four meter transition zone was required for the total accuracy to reach 75%, which is above the minimum threshold suggested by Thomlinson *et al.* (1999) but still below the target of 85% total accuracy. Such results suggest that the API errors observed are not just due to ambiguous plots which lie on the boundary between classes.

Non-stereo Photogrammetry

The general short coming of the API's ability to classify height is not unexpected as the use of non-stereo photogrammetric methods for predicting height, such as those used by the API, are known to be error prone even when classifying forests into very broad height categories such as the three used in the API (Wulder *et al.* 2010). The use of photointerpretation which relies on 2D/ non-stereo images is error prone as these images provide little information regarding the 3D morphology of objects in the images and as such aspects of tree geometry, such as height, cannot be accurately identified (Gong, Sheng & Biging 2002). Instead, 2D images are typically used to provide information on attributes (such as species) which are not reliant on tree geometry, since in many instances they are the only source of data available. The photo interpretation of height can be made accurate by utilising stereo pairs of images, although this technique relies on being able to measure from the base of the tree and so can only be used in open forest or on very flat ground (Véga & St-Onge 2008). Consequently, new techniques are being applied such as the use of accurate LiDAR (light detection and ranging) based approaches.

LiDAR

LiDAR uses the time elapsed between emission and detection of a laser pulse to measure the distance between a sensor and target (Baltsavias 1999). Given its ability to directly measure the structure of the vegetation as well as the underlying terrain, LiDAR is a valuable and precise tool for measuring vegetation height (with typical errors ranging from 0.5- 1.0m) in forests (Wulder *et al.* 2012, Véga & St-Onge 2008). Whilst these benefits have been demonstrated by numerous studies, to date very few large scale forest inventories have used this technology due to its high cost and lack of coverage (Tsui *et al.* 2012, Wulder *et al.* 2010). These limitations are likely to become less significant as new satellite missions become available, driven for example by the need to predict carbon stock in forests (e.g., LiDAR instrument on DESDynI, Goetz *et al.* 2009). Once LIDAR becomes more commonplace and economically feasible, it could be used to improve the accurate classification of forest height within API approaches. Furthermore, as a number of LiDAR based data sets already exist for parts of Victoria, these could be used to test the ability of LiDAR to improve the classification of forest height by the API.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

Ecoregions and the spatial variation in error

Spatial variation in the accuracy with which forest canopy height was classified was also observed. Ecoregion 1 obtained a higher total accuracy than the other two ecoregions. The higher accuracy in Ecoregion 1 could be due to the small range of tree heights typically encountered in this ecoregion. Heights in this ecoregion are generally lower, and there are no plots which fall in the high height class. The occurrence of less height classes within this ecoregion may have made the classification easier and produced more accurate results. In particular no medium height plots would be misclassified into the high class, which was a relatively common error that occurred in 12 out of 25 medium plots in Ecoregion 2, and in 21 out of the 53 plots in Ecoregion 3. The reference data for Ecoregion 3, similarly to Ecoregion1, only contained plots in the low and medium height class. However, operators expected trees to potentially fall in the high class in this ecoregion. As a direct consequence, misclassification errors between the medium and high classes could still occur. The misclassification of landscape elements between the medium and high class could be responsible for the spatial variation in error we have observed between ecoregions, and investigation into ways to eliminate or reduce this error would improve the API. Nevertheless, the accuracy of the API in classifying the height class of Ecoregion 1 was not optimal, therefore there must be other explanations for the general short comings of the API in classifying height.

The API was not as precise in classifying forest height as the two other variables previously analysed. It has been demonstrated that this error is significant and not just due to ambiguous plots which could belong to multiple classes. Equally, results indicate that the medium height class warrants investigation as the misclassification of these plots into the high height class could be responsible for the spatial variation observed between ecoregions. Finding ways to improve the classification of plots in the medium height class could, it is proposed, improve the API. Finally, as the use of non-stereo, on-screen photogrammetry to predict heights is known to be error prone the inclusion of supplementary data, such as the new highly accurate LiDAR based technology, could provide a way to improve the classification of the height class.

5 Conclusion

In summary, the API accurately classified forest/ non-forest land cover and vegetation type. This is excellent given the classification of forest/ non-forest land cover forms the foundation of all other classifications. The majority of errors in the classification of vegetation type occurred between mixed stands, that is stands comprised of *Eucalypt sp*. and one or more species, and pure *Eucalypt sp.* stands, both of which are similar vegetation classes. The classification of height was less precise, with the Mediterranean forest, woodland and scrub ecoregion obtaining more successful results. Height is known to be difficult to predict using non-stereo, on-screen photogrammetry alone and recent studies have shown the benefits of using LiDAR to not only record height accurately but estimate biomass which enables the prediction of carbon sequestration (Wulder *et al.* 2010). Given Victoria's forests store an estimated 511 million tonnes of carbon (DSE 2009), the accurate monitoring of carbon sequestration is very important. LiDAR may also provide the interpreters with a more detailed description of the ground cover in areas with low or sparse forest cover aiding in the classification of forest/ non-forest land cover and thus increasing the accuracy of this classification which is a starting point for all other classifications.

Future Directions

Three possible directions for future accuracy assessments are summarised below. The first is to rerun this analysis with the additional field plots to see if the same results are obtained, particularly in classes that lacked a sufficient amount of data to report users and producers accuracy, such as with the non-eucalypt vegetation types. The second is to analyse the API's accuracy in classifying canopy cover since this was not assessed in this study. And a third direction would be to investigate if the year of the image which was interpreted affects the accuracy of the results as the images were collected from 2006 to 2010.

Whilst there are still aspects of the accuracy to analyse and the classification of height can be improved with the inclusion of LiDAR information, this study has found the API to be accurate in classifying forest/ non-forest land cover and vegetation type. As such the API will aid in the reporting of the state, extent and condition of Victoria's public land forests by the DSE. This knowledge can in turn lead to informed management decisions, ultimately improving the sustainable management of these forests and ensuring that Victoria's environment is protected for future generations.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

References

Baltsavias, E. P., 1999. Airborne laser scanning: Basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, pp. 199–214.

Chapin, F III., Zavaleta, E., Eviner, V. Naylor, R., Vitousek, P., Reynolds, H., Hooper, D., Lavorel, S., Sala, O., Hobbie, S., Mack, M., & Diaz, S. 2000. Consequences of changing biodiversity. *Nature,* 405, pp. 234-242.

Congalton, R. 1988. Using spatial auto-correlation analysis to explore the errors in maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 54, pp. 587–592.

Congalton, R. & Green, K. 2009. *Assessing the accuracy of remotely sensed data, 2nd ed*. CRC Press, Boca Raton, pp. 105-140, 159.

Department of Sustainability and the Environment (DSE). 2007. Criteria and Indicators for Sustainable Forest Management in Victoria, Guidance Document. Victorian Government Department of Sustainability and Environment, Melbourne.

Department of Sustainability and the Environment (DSE). 2009. Victoria's state of the forests report 2008. Victorian Government Department of Sustainability and Environment, Melbourne.

Department of Sustainability and the Environment (DSE). 2011. Standard operating procedure 13, measuring a large tree plot. Victorian Government Department of Sustainability and Environment, Melbourne.

Farmer, E., Jones, S., Clarke, M., Buxton, L., Soto-Berelov, M., Page, S., Mellor, A. & Haywood, A. 2011. Semi-automated API for large area public land monitoring and reporting. Unpublished.

Foody, G. 1988. The effects of viewing geometry on image classification. *International Journal of Remote Sensing*, 9, pp. 1909–1915.

Foody, G. 2002, Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, pp. 185–201.

Foody, G. 2005. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *International Journal of Remote Sensing*, 26, pp. 1217-1228.

Goetz, S.J. Baccini, A., Laporte, N. T., Johns, T., Walker, W., Kellndorfer, J., Houghton, R.A., and Sun, M. 2009. Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. *Carbon Balance and Management*, 4, pp. 2.

Gong, P. Sheng, Y. & Biging, G.S. 2002. 3D model-based tree measurement from high-resolution aerial imagery. *Photogrammetric Engineering and Remote Sensing*. 68, pp. 1203- 1212.

Lillesand, T., Kiefer, R. & Chipman, J. 2004. *Remote sensing and image interpretation*, John Wiley & Sons, New York, pp. 215.

Linke, J., Betts, M., Lavigne, M. & Franklin, S. 2006. Structure, function, and change of forest landscapes. In: *Understanding forest disturbance and spatial pattern: remote sensing and GIS approaches*, CRC Press, Boca Raton, pp. 1-30.

Lucas, R., Cronin, N., Moghaddam, M., Lee, A., Armstrong, J., Bunting, P. & Witte, C. 2006. Integration of radar and Landsat-derived foliage projected cover for woody regrowth mapping, Queensland, Australia. *Remote Sensing of Environment*, 100, pp. 388 – 406.

McGwire, K. & Fisher, P. 2001. Spatially variable thematic accuracy: Beyond the confusion matrix. In: *Spatial uncertainty in ecology: Implications for remote sensing and GIS applications*. Springer, New York, pp. 308-329.

Morgan, J., Gergel, S. & Coops, N. 2010. Aerial photography: a rapidly evolving tool for ecological management. *Bioscience*, 60, pp. 47-59.

Olson, D. & Dierstein, E. 1998. The global 200: a representation approach to conserving the Earth's most biologically valuable ecoregions. *Conservation,* 12, pp. 502-515.

Pontius, R J. Jr. & Millones, M.2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32, pp. 4407- 4429.

Siry, J., Cubbage, F. & Ahmed, M. 2005. Sustainable forest management: Global trends and opportunities. *Forest Policy and Economics*, 7, pp 551- 561.

Steele, B., Winne, J. & Redmond, R. 1998. Estimation and mapping of misclassification probabilities for thematic land cover maps. *Remote Sensing of Environment*, 66, pp. 192–202.

Stehman, S. 2004. A critical evaluation of the normalized error matrix in map accuracy assessment. *Photogrammetric engineering and remote sensing*. 70, pp. 743- 751.

Stehman, S. & Foody, G. 2009. Accuracy assessment. In: *The SAGE handbook of remote sensing*. SAGE, London, pp. 297-309.

Thomlinson, J., Bolstad, P. & Cohen, W. 1999. Coordinating methodologies for scaling landcover classification from site-specific to global: steps towards validating global map products. *Remote Sensing of the Environment*,70, 16-28.

Tsui, O., Coops, N., Wulder, M., Marshall, P. & McCardle, A. 2012. Using multi-frequency radar and discrete-return LiDAR measurements to estimate above-ground biomass and biomass components in a coastal temperate forest. *ISPRS Journal of Photogrammetry and Remote Sensing*, 69, pp.121–133.

Véga, C. & St-Onge, B. 2008. Height growth reconstruction of a boreal forest canopy over a period of 58 years using a combination of photogrammetric and lidar models. *Remote Sensing of Environment*, 112, pp.1784–1794.

Vieira, C. & Mather, P. 2000. Visulisation of measures of classifier reliability and error in remote sensing. *Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Delft University Press, Delft, pp. 701-708.

Wulder, M., White, J., Stinson, G., Hilker, T., Kurz, W., Coops, N. & Benôit St-Onge, J.A. 2010. Implications of differing input data sources and approaches upon forest carbon stock estimation. *Environmental Monitoring & Assessment*, 166, pp.543–561.

Wulder, M., White, J., Nelson, R., Næsset, E., Ole Ørka, H., Coops, N., Hilker, T., Bater, C. & Gobakken, T. 2012. Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment*, 121, pp. 196–209

http://www.environment.gov.au/parks/nrs/science/bioregion-framework/terrestrial-habitats.html (2 Dec. 2012).

Acknowledgements:
We would like to thank the Department of Sustainability and the Environment and the University of RMIT, Melbourne, in particular my supervisors Simon Jones and Mariela Soto-Berelov for their guidance and assistance during the creation of this report. This study would not have been possible without the dataset provided by DSE. I would also like to thank the two anonymous reviewers for their valuable insight and constructive criticism.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5