Differentiation of ecosystems in the outback of Australia using airborne LiDAR

Auteur : Wagelmans, Isabelle
Promoteur(s) : Bastin, Jean-François; 21332
Faculté : Gembloux Agro-Bio Tech (GxABT)
Diplôme : Master en bioingénieur : sciences et technologies de l'environnement, à finalité spécialisée
Année académique : 2022-2023
URI/URL : http://hdl.handle.net/2268.2/18217

LIÈGE université
Gembloux
Agro-Bio Tech

Differentiation of ecosystems in the outback of Australia using airborne LiDAR

Isabelle Wagelmans

TRAVAIL DE FIN D'ETUDES PRESENTE EN VUE DE L'OBTENTION DU DIPLOME DE MASTER BIOINGENIEUR EN SCIENCES ET TECHNOLOGIES DE L'ENVIRONNEMENT

ANNEE ACADEMIQUE 2022-2023

CO-PROMOTEURS : Jean-François Bastin & Ben Sparrow

LIÈGE université
Gembloux
Agro-Bio Tech

Differentiation of ecosystems in the outback of Australia using airborne LiDAR

Isabelle Wagelmans

TRAVAIL DE FIN D'ETUDES PRESENTE EN VUE DE L'OBTENTION DU DIPLOME DE MASTER BIOINGENIEUR EN SCIENCES ET TECHNOLOGIES DE L'ENVIRONNEMENT

ANNEE ACADEMIQUE 2022-2023

CO-PROMOTEURS : Jean-François Bastin & Ben Sparrow

Acknowledgment

First and foremost, this paper and the research behind it would not have been possible without the support of my supervisor, Jean-François Bastin. Thank you for this amazing subject, for your guidance, your very understandable explanations and for your kindness.

I also want to express all my gratitude to Ben Sparrow who made this thesis feasable. Thank you for this incredible opportunity, the unforgettable trip to the Australian Outback and for your kindness.

This work is supported by the use of Terrestrial Ecosystem Research Network (TERN) infrastructure, which is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS).

This project would not have been possible without the financial support of the Erasmus Plus program. I also want to thanks the Fédération Wallonie-Bruxelles and the Belgian Welfare without whom I would not have been able to even start studying.

I would like to give a big thanks to the entire TERN team who welcomed me so well, facilitated my adjustment at the other side of the world and made my stay enjoyable.

I take this opportunity to thanks Poornima Sivanandam, from the University of Tasmania, without whom the processing of raw LiDAR data would have been very tedious.

I greatly appreciate the help from Joelle Bezzan and Pauline Depoortere for their constructive feedback on my writings.

Of course, I want to especially thanks Kim Hoang, aka Jesus, aka The Savior. Once again, you demonstrated that you live up to your nickname. Thank you for your last minute help, your encouragement and your kindness. You're a real one !

Also, a big shout out to Australia for sparing me. We both know what you are capable of and I appreciate that you only showed me your good side.

Last but not least, I am grateful for the support of my family. Thank you for putting up with me. It has been a hell of a few months!

Abstract

Monitoring ecosystems plays a key role in facing climate change impacts. A better understanding of ecosystem functioning is needed to take appropriate actions toward their conservation. Ecosystems of the outback of Australia are of major importance because of their endemism, their vital services to local populations and their fragility caused by human pressures and climate change. Remote sensing is widely used for large-scale ecosystem monitoring however, drylands remote sensing faces unique challenges not typically encountered in other regions. Their high heterogeneity and high soil background reflectance are just a few of the difficulties encountered in these lands. The innovative LiDAR technology has the advantage of detecting the three-dimensional structure of the vegetation and could potentially overcome the uncertainty generated by optical imaging in arid and semi-arid areas.

The general goal of this work is to study the potential use of high resolution airborne LiDAR data to classify different ecosystems found in the outback of South Australia. In order to characterize the ecosystems, various structural components were calculated from LiDAR point clouds. The relevance of these metrics in the discrimination of our ecosystems was assessed through a PCA coupled with an analysis of their descriptive statistics. Three classification models were build (a hierarchical clustering, a decision trees and a LDA) with different numbers of input variables. Their performance was compared with the accuracy related to the confusion matrix. The two-variable, top of canopy height and number of trees, models offer the best compromise between parsimony and accuracy. The LDA is the best predictor with an accuracy of 84%. However, the decision tree, whose overall accuracy is only 1% less, is easier to interpret and to relate to ecological reality.

Then, in order to evaluate the ability of the models to be upscaled, discriminant models were tested on larger plots, but the results were not satisfactory. This is due to the number of trees, used as input variable, being dependent on the size of the plot. The use of GEDI data was also explored to assess the potential of models to be extrapolated to the global scale. However, the comparison of airborne and spaceborne LiDAR metrics revealed a significant difference between the two datasets.

The results confirm the hypothesis that metrics derived from high resolution airborne LiDAR are capable of discriminating some Australian ecosystems. Nonetheless, this study is a first approach and further research are required to improve the large-scale characterization of drylands ecosystems.

Résumé

La surveillance des écosystèmes joue un rôle essentiel dans la lutte contre les effets du changement climatique. Une meilleure compréhension du fonctionnement des écosystèmes est nécessaire pour prendre les mesures appropriées en vue de leur protection. Les écosystèmes de l'outback australien revêtent une grande importance en raison de leur endémisme, des services écosystémiques rendus aux populations et de leur vulnérabilité aux pressions anthropiques et au changement climatique. La télédétection est largement utilisée pour la surveillance à large échelle des écosystèmes toutefois, la télédétection dans les terres arides se heurte à des difficultés particulières que l'on ne rencontre généralement pas dans d'autres régions. Leur grande hétérogénéité et la réflectance élevée du sol ne sont que quelques unes des difficultés spécifiques à ces zones. La technologie innovante LiDAR a l'avantage de détecter la structure tridimensionnelle de la végétation et pourrait potentiellement surmonter l'incertitude générée par l'imagerie optique dans les régions arides et semi-arides.

Le but général de ce travail est d'étudier l'utilisation potentielle de données LiDAR aérien à haute résolution pour classer différents écosystèmes de l'outback de l'Australie méridionale. Afin de caractériser les écosystèmes, diverses composantes structurales ont été calculées à partir de nuages de points LiDAR. La pertinence de ces paramètres pour la discrimination de nos écosystèmes a été évaluée à l'aide d'une ACP couplée à une analyse de leurs statistiques descriptives. Trois modèles de classification ont été construits (un regroupement hiérarchique, un arbre de décision et une ADL) avec differents nombres de variables d'entrée. Leur efficacité a été comparée grace à la précision globale associée à la matrice de confusion. Les modèles à deux variables, à savoir, la hauteur de canopée et le nombre d'arbres, offrent le meilleur compromis entre parcimonie et précision. La ADL est le meilleur prédicteur avec une précision de 84%. Cependant, l'arbre de décision, dont la précision globale n'est que de 1% inférieure, est plus facile à interpréter et à relier à la réalité écologique.

Ensuite, afin d'évaluer la possibilité de transposer les modèles à plus grande échelle, les modèles discriminants ont été testés sur de plus grandes parcelles, mais les résultats n'ont pas été satisfaisants. Cela est dû au fait que le nombre d'arbres, utilisé comme variable d'entrée, dépend de la taille de la parcelle. L'utilisation des données GEDI a également été étudiée pour évaluer le potentiel d'extrapolation des modèles à l'échelle mondiale. Cependant, la comparaison des métriques LiDAR aérien et spatial a révélé une différence significative entre les deux jeux de données.

Les résultats confirment l'hypothèse selon laquelle les métriques dérivées du LiDAR aérien à haute résolution sont capables de différencier certains écosystèmes australiens. Néanmoins, cette étude n'est qu'une première approche et d'autres recherches sont nécessaires pour améliorer la caractérisation à grande échelle des écosystèmes des terres arides.

Contents

1 Introduction **7**
- 1.1 Key role of monitoring ecosystems in facing climate change impacts . . . 7
- 1.2 Uniqueness of Australian ecosystems and particularly drylands . . . 7
- 1.3 Challenges in drylands remote sensing . . . 10
- 1.4 Potential of LiDAR to monitor ecosystems . . . 10
- 1.5 Research objectives . . . 12

2 Materials and methods **13**
- 2.1 Study area . . . 13
- 2.2 Data collection and pre-processing . . . 17
- 2.3 Data processing . . . 17
  - 2.3.1 Canopy Height Model . . . 17
  - 2.3.2 Metrics . . . 18
- 2.4 Principal Component Analysis . . . 20
- 2.5 Classification models . . . 20
- 2.6 Upscaling . . . 21
- 2.7 GEDI . . . 21

3 Results **22**
- 3.1 Descriptive statistics on the variables . . . 22
- 3.2 Principal Component Analysis . . . 24
- 3.3 Classification models . . . 24
  - 3.3.1 Comparison of models with different numbers of variables . . . 24
  - 3.3.2 Two variables models . . . 25
- 3.4 Upscaling . . . 27
- 3.5 GEDI . . . 28

4 Discussion **30**
- 4.1 Variables analysis . . . 30
- 4.2 Classification models . . . 30
  - 4.2.1 Comparison of models with different numbers of variables . . . 30
  - 4.2.2 Two variables models . . . 31
  - 4.2.3 Limitations . . . 32
- 4.3 Upscaling . . . 33
- 4.4 GEDI . . . 33
- 4.5 Additional prospects . . . 34

5 Conclusion **35**

6 Contribution of the student **36**

7 References **37**

8 Appendices **43**

List of Figures

1 Chenopod shrubland panorama. 8
2 Mallee panorama. 8
3 Mulga panorama. 9
4 Eucalypt woodland panorama. 9
5 Distribution of major vegetation types in Australia. Locations of Calperum Reserve and Wintinna Station are represented by black squares. Map was generated based on Australia's National Vegetation Information System—Major Vegetation Groups (NVIS-MVGs). Groups were obtained by reclassifying the original 26 NVIS-MVGs. Reproduced from Eamus et al. (2016) 14
6 Division of a plot of 100m side into sub-plots of 25m side. 15
7 Google Earth images of the plots. Each row represents an ecosystem. In order: mulga, mallee, chenopod shrubland, eucalypt woodland. (a) saagvd0005, (b) saagvd0007, (c) saagvd0008, (d) sasmdd0001, (e) sasmdd0002, (f) sasmdd0003, (g) sasmdd0005, (h) sasmdd0011, (i)saastp0036, (j)sasmdd0008, (k) sasmdd0012, (l) sasmdd0013. 16
8 Descriptive statistics on the variables. For each metric, a comparison of the four ecosystems is made on the basis of boxplots. Boxplots include the data median, range and dispersion as well as outliers. The x axis represents the ecosystems: chenopods, eucalypt, mallee and mulga. The y axis represents each variable: top of canopy height, maximum height, standard deviation height, return frequency between 0 and 2m, return frequency between 2 and 4m, return frequency between 4 and 6m, return frequency between 6 and 8m, return frequency between 8 and 10m, return frequency above 10m, canopy density, canopy cover, number of trees, mean crown area and total crown area. 23
9 Plots classification using hierarchical clustering represented in the first factorial plane of the PCA. The x and y axis are the first and second dimension of the PCA respectively. Each dot represents a sub-plot. Each color is associated with an ecosystem: chenopods is blue, mulga is red, mallee is green and eucalypt is purple. 25
10 Decision tree classification. Dichotomous key applying thresholds on the two variables, the top of canopy height (chm_mean) and the number of trees (ntree), to create groups. Each color represents an ecosystem: chenopods are red, eucalypt is orange, mallee is grey and mulga is green. 26
11 Comparison between the maximum height calculated from drone (M300) data and from GEDI data. Boxplots include the data median, range and dispersion as well as outliers. The x axis represents the ecosystems: chenopods, eucalypt, mallee and mulga. The y axis is the maximum elevation (m). The red boxplots are associated with GEDI and the blue ones with the drone (M300). 28
12 Comparison between the canopy cover calculated from drone (M300) data and from GEDI data. Boxplots include the data median, range and dispersion as well as outliers. The x axis represents the ecosystems: chenopods, eucalypt, mallee and mulga. The y axis is the canopy cover (-). The red boxplots are associated with GEDI and the blue ones with the drone (M300). 29
13 Point cloud of sasmdd0008 constructed using the DJI Terra software. 43

List of Tables

1 Computed metrics, their type and their unit. 19

2 Comparison of models with different numbers of variables. Accuracies of the confusion matrices of the three models (hierarchical clustering, decision trees and LDA) with 14, 4 or 2 variables as input. The best model seems to be the LDA with 14 input variables. . . 24

3 Performance criteria of the hierarchical clustering. Sensitivities and specificities associated with each ecosystem (mulga, mallee, chenopods and eucalypt) as well as the accuracy of the confusion matrix for the hierarchical clustering classification. The overall accuracy is 80%. 25

4 Performance criteria of the decision tree classification. Sensitivities and specificities associated with each ecosystem (mulga, mallee, chenopods and eucalypt) as well as the accuracy of the confusion matrix for the decision tree classification. The overall accuracy is 83%. 26

5 Performance criteria of the LDA classification. Sensitivities and specificities associated with each ecosystem (mulga, mallee, chenopods and eucalypt) as well as the accuracy of the confusion matrix for the LDA classification. The overall accuracy is 84%. 27

6 Comparison of models accuracy tested with various plot sizes. Accuracies of the discriminant models (decision tree and LDA) trained using 25m square plots and tested with 50m or 100m square plots. 27

7 Site name, latitude and longitude of the up right corner of the plot, location and ecosystem of each site. 43

1 Introduction

1.1 Key role of monitoring ecosystems in facing climate change impacts

Monitoring ecosystems is the first step towards their conservation. In fact, we need to better understand the functioning of ecosystems and consequently, the impact of a changing environment on our natural resources in order to ensure sustainable conservation and management strategies (Sparrow et al., 2020). Monitoring allows us to better understand ecosystems and to take appropriate actions to maintain their integrity.

One of the key goals for ecosystems monitoring is biodiversity conservation. Indeed, understanding species diversity and monitoring populations are fundamental for the conservation of threatened species. Another important aspect of ecosystems monitoring is the detection of climate change effects. Ecosystems are sensitive to climate change and their monitoring detects climate-related changes and impacts, such as temperature variations, precipitation, extreme weather events, and species migrations. This information is essential for assessing the magnitude of climate change and for developing adaptation strategies.

1.2 Uniqueness of Australian ecosystems and particularly drylands

Monitoring Australia's ecosystems is especially important because of its unique biodiversity. Indeed, Australia is home to an exceptional array of flora and fauna. Many Australian species are endemic which means that they cannot be found anywhere else in the world. What is more, Australian ecosystems are often characterized by extreme environmental conditions, including arid climates, drought cycles, frequent fires and nutrient-poor soils. These factors make Australian ecosystems particularly sensitive to environmental disturbances and changes. 70% of Australia is either arid or semi arid land; it is the driest inhabited continent in the world [1].

Drylands cover about 40% of the Earth's surface, and billions of people depend on these vital ecosystem services (Adeel, 2005). Arid regions are highly climate sensitive and respond strongly to rainfall changes. More than a third of the world's biodiversity hotspots are located in drylands (Myers et al., 2000). Besides, the global area of drylands is increasing rapidly. This expansion will lead to reduced carbon sequestration and increased regional warming. Those factors coupled with the rapid human population growth will exacerbate the risk of land degradation and desertification in the near future (Huang et al., 2016). Monitoring dryland ecosystems is therefore particularly relevant.

[1] https://www.dcceew.gov.au/environment/land/rangelands

Ecosystems studied

For the feasibility of this project, only four vegetation types were selected among Australia's great biodiversity. The ecosystems concerned in this study are characterized by endemic species and a semi-arid climate.

i. Chenopod shrubland

Chenopod shrublands are dominated by various hardy low shrub species typically belonging to the Chenopodiaceae family. Scattered emergent trees and shrubs are sometimes present. They are mostly distributed in southern arid and semi-arid rangelands on extensive clay plains, undulating gibber hills and plains and aeolian sandplains. They are widely affected to livestock grazing ("NVIS Fact sheet MVG 22 – Chenopod shrublands, samphire shrublands and forblands", 2017). They represent about 6% of the total land area of mainland Australia. They are restricted to dry climates where mean annual rainfall is less than 350 mm, of which up to 50% falls in winter. Chenopods are characterized by low transpiration rate, high water use efficiency and their resistance to drought and salinity (Foulkes et al., 2014).

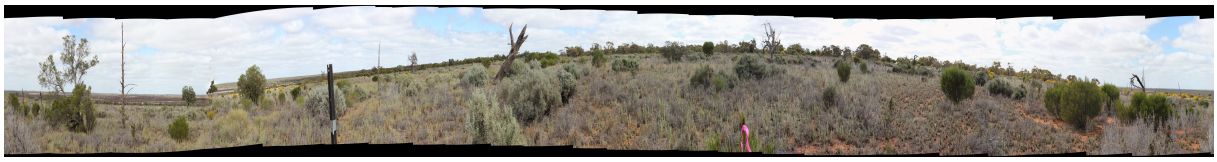

Figure 1: Chenopod shrubland panorama.

ii. Mallee

Mallee is one of the dominant plant communities of southern Australia. It refers to vegetation dominated by multi-stemmed eucalypts growing as tall shrubs or short trees (2-10m). It is found in the transitional zone between southern temperate sub-humid woodlands and northern arid flora (Peart, 1990). It typically grows under semi-arid Mediterranean climate with cool winters and hot dry summers. Rainfall occurs predominately in winter. Mallee eucalypts are sclerophyllous and produce multiple stems from a rootstock known as a lignotuber which is an adaptation to wildfire. The understory composition is various and depends on local weather conditions. Mallees have a rich and unique biodiversity. Indeed, many species of plants, birds, reptiles and invertebrates cannot be found in any other ecosystems ("NVIS Fact sheet MVG 32 – Mallee open woodlands and sparse mallee shrublands", 2017).

Figure 2: Mallee panorama.

iii. Mulga

Mulga describes both the name of a plant (*Acacia aneura* F. Muell. ex Benth and its close relatives), and the name of a vegetation type. Across the country, forms of mulga are highly variable, ranging from relatively short (<3m) to relatively tall (>10m) trees, and ranging from a very open canopy to a relatively closed canopy. This vegetation type covers about 20% of the land surface of Australia (Eamus et al., 2016). It is most commonly found on red-earth soils. Climatic conditions are generally hot summers with a pronounced dry season and cool to warm winters. A sparse small shrub layer may occur and ephemeral herbs and plants may cover the ground in response to rain ("NVIS Fact sheet MVG 16 – Acacia shrublands", 2017).

Figure 3: Mulga panorama.

iv. Eucalypt woodland

Eucalypt woodlands are the characteristic vegetation in flat to gently hilly landscapes of the intermediate rainfall zones of South Australia (Keith, 2017). They form a transitional zone between the high rainfall forests on the margins of the continent and the hummock grasslands and shrublands of the arid interior. They are usually composed of scattered eucalypts and bloodwoods (*Corymbia*) over an understory of grasses and sparse shrubs. Trees are up to 30m high ("NVIS Fact sheet MVG 5 – Eucalypt woodlands", 2017).

Figure 4: Eucalypt woodland panorama.

1.3 Challenges in drylands remote sensing

Remote sensing is a key technology for large-scale vegetation monitoring (Cabello et al., 2012; Nagendra et al., 2013). Satellite remote sensing measurements are widely accessible, they offer a cheap and verifiable means to obtain complete spatial coverage of environmental information (Pettorelli et al., 2014). However, rapid innovations in drones, cameras and 3D photogrammetry have made drone remote sensing a good alternative. Drone-based products can accurately and efficiently provide ultra-high resolution imagery at relatively large scale (J. Zhang et al., 2021). The benefits of drone remote sensing in image acquisition include high spatial and temporal resolution, full flexibility, low material and operational costs. Another advantage is that Unmanned Aerial Vehicles (UAV), also known as drones, overcome the drawback of the ground based system such as the inaccessibility to muddy or very dense regions (Bansod et al., 2017).

Unfortunately, dryland remote sensing faces unique challenges not typically encountered in mesic or humid regions. The major challenge is the high heterogeneity of drylands at many scales, which is linked to the various vegetation forms, structures and functions (e.g., evergreen and deciduous shrubs) (Ganem et al., 2022). Other difficulties faced in drylands include low vegetation signal-to-noise ratios, high soil background reflectance and irregular growing seasons due to unpredictable seasonal rainfall and frequent periods of drought (Smith et al., 2019).

As a result, different studies monitoring vegetation in drylands do not systematically produce the same results. For instance, significant spatial disagreements are observed on two satellite-based global forest maps (Hansen et al., 2003; Sexton et al., 2013). Tree cover mapping is known to produce very uncertain results in drylands (Sexton et al., 2016). Ko et al. (2009) also agreed that quantifying tree canopy cover in dryland ecosystems from remote sensing data is particularly difficult due to low tree cover and density, short tree stature, and co-existence of trees and shrubs at fine spatial scales.

What is more, our knowledge of the extent of vegetation in these ecosystems is limited (Bastin et al., 2017). There is still a lack of studies assessing dryland biomes (Durant et al., 2012; Ruusa et al., 2022) as well as ground-based data needed for calibration and validation of remote sensing algorithms (Smith et al., 2019).

1.4 Potential of LiDAR to monitor ecosystems

LiDAR (standing for Light Detection And Ranging) appears to be the most promising sensor for describing vegetation structure (Huylenbroeck et al., 2020). LiDAR sensor measures the time between the emitted infrared light pulse and the returned pulse. Knowing the speed of light, this return time allows to calculate the precise distance between the sensor and the object on which the light beam is reflected. The processed LiDAR result is known as a point cloud. Point clouds are large collections of 3D elevation points, which include x, y, and z, along with additional attributes.

The use of LiDAR has the advantage of detecting the three-dimensional structure of the ecosystem. In other words, this technology is able to provide structural information on both the vertical and the horizontal axis (Lim et al., 2003; Michez et al., 2016). Indeed, LiDAR can measure various biophysical components such as canopy height, individual tree height, vertical distribution or crown diameter (Dubayah & Drake, 2000; Lim et al., 2003; Popescu & Wynne, 2004; Popescu et al., 2002). Moreover, LiDAR measurements penetrate dense forest vegetation to generate accurate estimates of surface topography and canopy height (Leitold et al., 2015).

Two main approaches exist to calculate LiDAR-derived metrics: area-based and tree-centric. The first method is mainly used for regional or national scales. It relates estimates obtained from reference field plots to simple metrics, often, linked to canopy height (Coomes et al., 2017). Some of the most commonly calculated variables are the mean canopy height and the top of canopy height (Asner et al., 2012; Meyer et al., 2013). The latter is usually preferred due to its better overall stability among the various LiDAR sensors, whether airborne or spaceborne (Asner & Mascaro, 2014). On the one hand, computation of those metrics is convenient. On the other hand, the associated statistics are simple even with low-resolution data. Therefore, this approach is prevalent across literature (Coomes et al., 2017).

The tree-centric approach identifies and modelizes each tree belonging to the plot. Their related metrics can then be integrated into allometric equations and summed to obtain plot-level estimates (Ferraz et al., 2016). To do so, Individual Tree Crowns (ITC) segmentation techniques detect canopy trees in the point clouds, giving their position, their individual height and providing an insight of the stem density in the area (Popescu et al., 2002). Satisfying a predefined algorithm, the surrounding points are classified as belonging to the tree crown (Dalponte & Coomes, 2016; Ferraz et al., 2016; Hyyppa et al., 2001). Then, tree crown area or even canopy cover can be retrieved from the resulting objects. Although most of the techniques used to extract individual trees focus on the upper layers and the canopy, recent progress has been made in developing algorithms capable of modelling every layer of the forest (Ferraz et al., 2016).

As mentioned by de Lame (2021), one of the biggest interests of these tree-centric methods lies in their similarity with field-based approaches allowing for the development of analogous allometric methods. The common theoretical basis facilitates the identification and understanding of uncertainties and biases. However, over- or under-segmentation of trees can lead to biases, the associated statistics are more complicated to manipulate, and the computation of these methods can be intensive. Advances made in individual tree crowns methods through the development of new algorithms and sensors may give this approach a more prominent role in the future of structural components estimation (Coomes et al., 2017).

A last important point to mention is the use of spaceborne LiDAR. In 2018, NASA launched the Global Earth Dynamics Investigation (GEDI) large-footprint sensor on the International Space Station. GEDI produces high resolution laser ranging observations of the 3D structure of the Earth with the aim of providing precise measurements of forest canopy height, canopy vertical structure, and surface elevation [2]. Theoretically, GEDI could be able to extract proxies of forest structure at the plot scale after a calibration phase from a comparison with airborne LiDAR data (Chave et al., 2019; Schneider et al., 2020). GEDI data promise to be particularly useful given that they are worldwide and accessible to everyone.

[2] https://gedi.umd.edu/mission/mission-overview/

1.5 Research objectives

As explained above, remote sensing in drylands generates a lot of uncertainties and so far, limited research has been engaged on this. Therefore, this study aims to improve the global knowledge on the subject. The general goal of this work is to study the potential use of high resolution airborne LiDAR data to classify different ecosystems (chenopod shrubland, mallee, mulga and eucalypt woodland) found in the outback of South Australia.

In order to characterize the ecosystems, various structural components will be calculated from LiDAR point clouds. The following step is to identify which of these structural metrics are the most relevant to discriminate the studied ecosystems. Descriptive statistics and Principal Component Analysis will help answer that question. Afterwards, using the LiDAR-derived variables as input, three different classification models will be developed and compared so as to select the most effective one. Finally, a comparison between metrics calculated from drone data and from GEDI satellite data will be carried out in order to evaluate the potential use of the selected model with large-scale data.

2 Materials and methods

2.1 Study area

This study is based in the Australian outback, and more precisely, in the state of South Australia. The data were collected on two different locations: the Calperum Reserve and the Wintinna Station (Figure 5). These two areas are separated by a distance of about 952km [3] as the crow flies and are described in more detail below.

The Australian Commonwealth Bureau of Meteorology divides Australia into 28 climate classes [4] using a modified Köppen climate classification scheme based on temperature as well as the amount and the timing of rainfall (Williams et al., 2012).

i Calperum Reserve
The Calperum Reserve is located in the Murray Darling Depression. According to the Australian Commonwealth Bureau of Meteorology, the climate of this region is a warm (persistently dry) grassland. Calperum is part of the winter dominant zone which means that is has a marked wet winter and a dry summer. The mean annual temperature is 17°C [5] and the mean annual rainfall is around 240mm [6].

ii Wintinna Station
The Wintinna Station is straddled between the Great Victoria Desert and the Stony Plains. As classified by the Australian Commonwealth Bureau of Meteorology, the climate is a hot (persistently dry) desert. It is part of the arid zone due to low rainfall. The mean annual temperature is 21°C [7] and the mean annual rainfall is around 150mm [8].

Figure 5, reproduced from Eamus et al. (2016) [TERN], shows the distribution of major vegetation types in Australia. The locations of Calperum Reserve and Wintinna Station were added for illustrative purposes. Mulga and shrubland can be found in the Wintinna Station. A mix of forest, savanna, shrubland and a little bit of mulga can be encountered in the Calperum Reserve.

[3] https://www.distancede.com/Distance-de-Vol-calculateur.aspx
[4] http://www.bom.gov.au/climate/maps/averages/climate-classification/?maptype=kpn
[5] http://www.bom.gov.au/climate/averages/tables/cw_024048.shtml
[6] http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_stn_num=024048
[7] http://www.bom.gov.au/climate/averages/tables/cw_016007.shtml
[8] http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_stn_num=016093

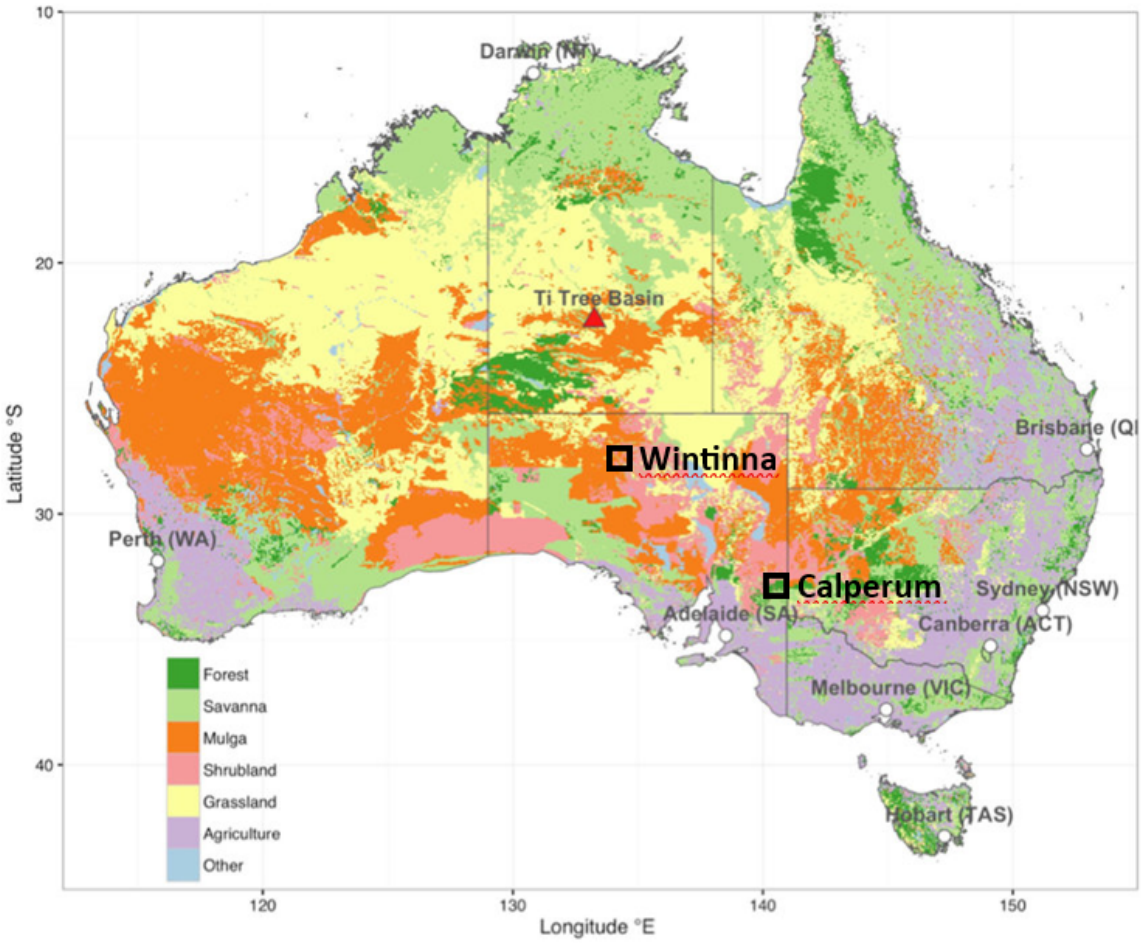

Figure 5: Distribution of major vegetation types in Australia. Locations of Calperum Reserve and Wintinna Station are represented by black squares. Map was generated based on Australia's National Vegetation Information System—Major Vegetation Groups (NVIS-MVGs). Groups were obtained by reclassifying the original 26 NVIS-MVGs. Reproduced from Eamus et al. (2016)

A plot is a 100m square with an area of 1ha. Each ecosystem is represented by three sites. In total, 12 sites were selected in this study. Figure 7 shows the Google Earth image of each site (each row represents an ecosystem: mulga, mallee, chenopod shrubland and eucalypt woodland respectively). In order to increase the amount of data, each site was divided into 16 sub-sites. Each sub-site is a square plot of 25m side (Figure 6). The Calperum data were collected in May 2022 and the Wintinna data were collected in October 2022. The GPS coordinates of each site as well as their location and the ecosystem they belong to can be found in Table 7 in the Appendices.

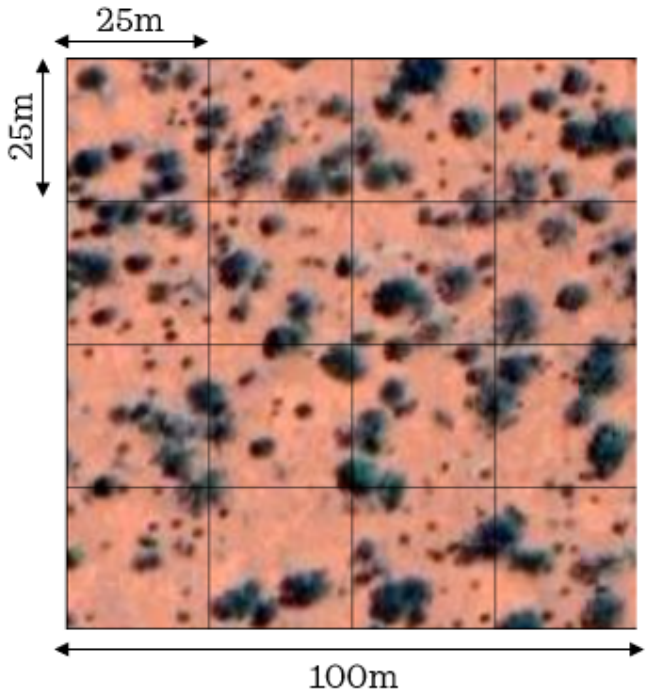

Figure 6: Division of a plot of 100m side into sub-plots of 25m side.

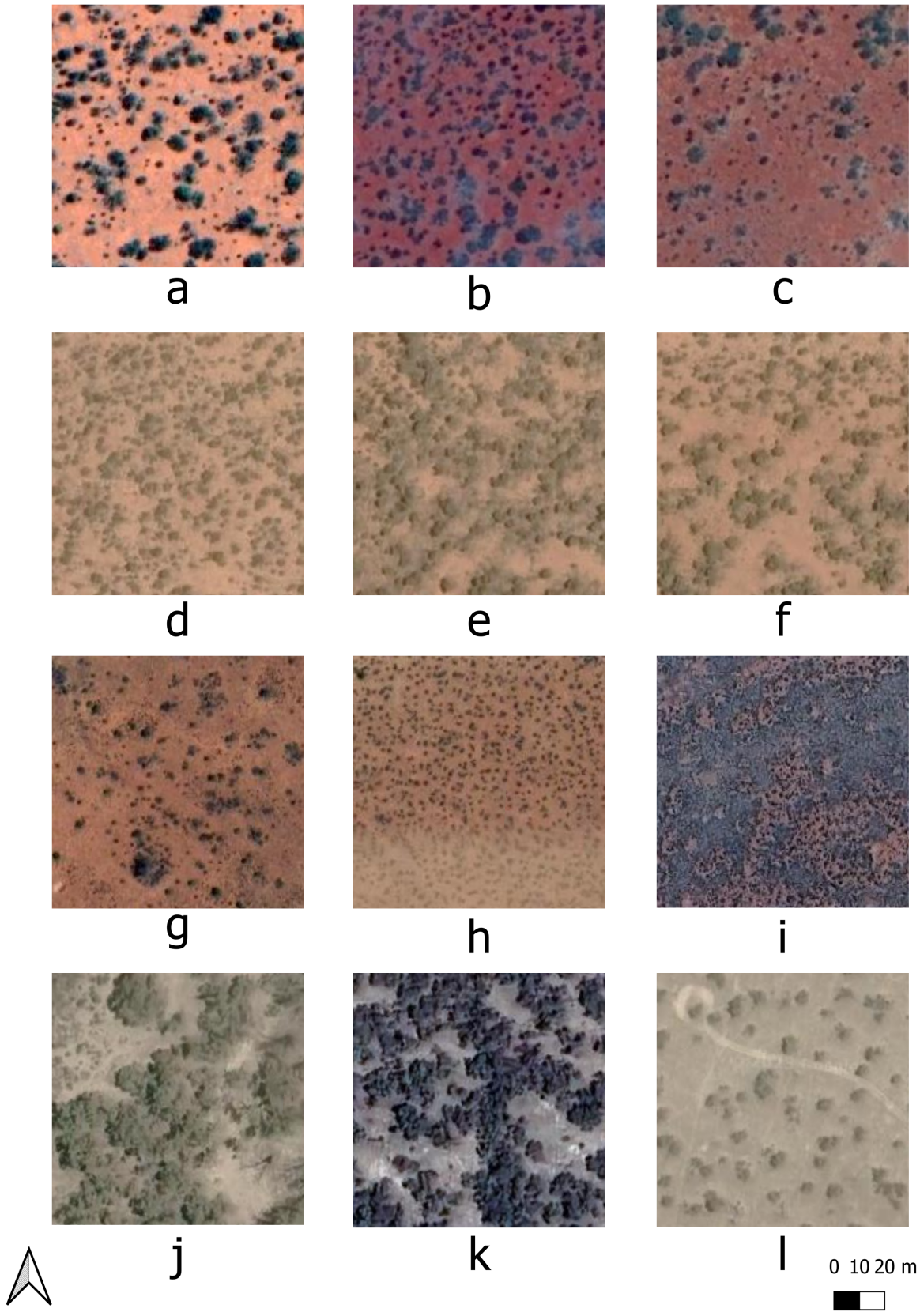

Figure 7: Google Earth images of the plots. Each row represents an ecosystem. In order: mulga, mallee, chenopod shrubland, eucalypt woodland. (a) saagvd0005, (b) saagvd0007, (c) saagvd0008, (d) sasmdd0001, (e) sasmdd0002, (f) sasmdd0003, (g) sasmdd0005, (h) sasmdd0011, (i)saastp0036, (j)sasmdd0008, (k) sasmdd0012, (l) sasmdd0013.

2.2 Data collection and pre-processing

Data were acquired by airborne LiDAR: flying a drone equipped with a LIDAR sensor. The Unmanned Aerial System (UAS) used for this study was the DJI Matrice 300 (M300) RTK. LiDAR data were collected using the DJI Zenmuse L1 sensor. The following settings were applied for the drone flights:

- Flight route altitude: 50m;
- Target Surface to Takeoff Point: 0 (for chenopod shrublands), 5 (for the others);
- Speed: 5m/s;
- Side overlap: 50%;
- Margin (buffer): 50m;
- Return Mode: Triple (for forests), Dual (for low-lying vegetation);
- Sampling Rate: 160kHz (for forests), 240kHz (for low-lying vegetation);
- Scanning Mode: Repetitive.

The point clouds were constructed using the DJI Terra software and the results were saved as .las files. Figure 13, found in the Appendices, is an example of the processed 3D image. Flying at low altitudes at relatively slow speeds produce point densities whose orders of magnitude are greater than traditional airborne laser scanning. Moreover, by operating at low altitude, errors in horizontal position of projected laser pulses are minimized (Kellner et al., 2019).

2.3 Data processing

2.3.1 Canopy Height Model

The first processing step was to separate the ground and non-ground (vegetation) points from the LiDAR point cloud. This ground classification was done using the Cloth Simulation Filter (CSF) algorithm. This method works on raw LiDAR data and can be applied to various landscapes without determining elaborate filtering parameters (W. Zhang et al., 2016).

After height normalization and filtration of the outliers, a Canopy Height Model (CHM) was generated using the p2r algorithm. This method densifies the point cloud and the resulting canopy model is smoother and contains fewer 'pits' and empty pixels (LidR documentation). The high density of the point cloud allows a high resolution (Höfle & Hollaus, 2010). The produced CHM has a 5cm resolution as required by the Terrestrial Ecosystem Research Network (TERN) procedure.

As a 50m buffer was set around the plots to avoid edge effects, the CHM needed to be cut into 100m square which are the initial plots. Only then are the final CHM divided into 25m sub-plots.

2.3.2 Metrics

Various metrics were computed in order to later determine which ones were best suited to discriminate the four ecosystems (chenopod shrubland, mallee, mulga and eucalypt woodland). Those variables are commonly found in the literature to characterize ecological structure (**to name just a few**; Ferraz et al., 2016; Lim et al., 2003; Meyer et al., 2013; Miura & Jones, 2010; Trouvé et al., 2023; Z. Zhang et al., 2011). They can be divided into three types: vertical, horizontal and tree-centric metrics. Table 1 is a list of all the variables and their unit.

Vertical metrics

Asner and Mascaro (2014) recommended using the top of canopy height (TCH) instead of the mean canopy height due to the sensor-specific variability of the latter. Indeed, vertical profile metrics calculated using the whole waveforms are highly sensitive to the LiDAR sensor. The TCH corresponds to the average value of the CHM pixels in a plot. Therefore, the four following metrics were calculated directly from the CHM.

The top of canopy height is computed as the average height of each pixel within a plot. The maximum height is determined as the maximum value of all the pixels inside a plot. The standard deviation height is the standard deviation of the pixel values belonging to a plot. The vertical profile is comprised of 6 sub-metrics which each represents the pixel frequency between two height thresholds (0-2m, 2-4m, 4-6m, 6-8m, 8-10m, 10m and above [as only the eucalypt woodlands reach higher than 10m]).

Horizontal metrics

Canopy density and canopy cover are both ratios of vegetation to ground as seen from the air. The first parameter is the fraction of the number of returns above a 1.4m threshold (defined by TERN) over the total number of returns. The second metric is similar to the first one, except that it only applies to the first return. It is therefore the ratio of the number of first returns above a 1.4m threshold (defined by TERN) over the total number of first returns.

Tree-centric metrics

The locate_trees algorithm in the lidR package was used to spot local maxima in the normalised point cloud (Roussel et al., 2020). To count the number of trees, it is assumed that each one of those maxima corresponds to a tree. In order to avoid the most omission and commission errors, it is important that the size of the measurement window considered by the algorithm is proportional to the area of the crown of the tree (Popescu & Wynne, 2004). Therefore, different window sizes were tested. Knowing that the crown area is directly proportional to the size of the tree, an algorithm using a moving measurement window (WS) which varies with the height (H) of the LiDAR point under consideration is pertinent. de Lame (2021) found that the equation (1) developed by Blanchard et al. (2016) was the most appropriate.

$$WS = \sqrt{\frac{0.225 \times H^{1.946}}{\pi}} \tag{1}$$

Unfortunately, this method turned out to largely overestimate the number of trees. A constant measurement window was therefore applied. A sensitivity analysis determined that a 5m window was the best fitted for this study. The validation was done through a supervised accuracy assessment (i.e., assessing the sensitivity of the algorithm by visually comparing the number of trees detected with the number of trees on the RGB image of the corresponding plot). The maximum heights and the geographical coordinates of the detected trees were saved in order to be used for the next metrics calculation.

Individual tree crowns (ITC) segmentation techniques allow to delineate tree crowns and thus, determine their areas. ITC segmentations require two steps: identifying trees and then segmenting crowns. The first step had already been done when computing the number of trees. For the second step, two techniques were carried out. Both methods are parametric algorithms using the LiDAR point cloud and the CHM to achieve ITC segmentation (Roussel et al., 2020). The first method was developed by Dalponte and Coomes (2016) and is an adaptation of the technique created by Hyyppa et al. (2001). The second method was developed by Silva et al. (2016). After a visual comparison of the final crowns segmentation with the RGB image of the plot, the Dalponte and Coomes (2016) algorithm seemed to underestimate the crowns area. Therefore, the Silva et al. (2016) method was kept to calculate the crowns area of each plot.

This technique applies a variable circular crown buffer to delimit the initial tree crown area. This parameter is user-defined and corresponds to a proportion of the tree height. A sensitivity analysis determined that 80% of the tree height was best fitted for this study. Next, the initial crown areas are split using the centroidal voronoi tessellation approach to isolate each individual tree polygon. Then, the grid cells which values are below 30% of the maximum height in each detected tree are excluded to eliminate the low-lying noise. Finally, the crown delineation is computed by delimiting the boundary of grid cells belonging to each tree (Silva et al., 2016). From this segmentation, the total crown area and the mean area of each plot were calculated.

Table 1: Computed metrics, their type and their unit.

Type	Metric		Unit
Vertical	Top of canopy height		m
	Maximum height		m
	Standard deviation height		m
	Vertical profile	Return frequency between 0 and 2m	%
		Return frequency between 2 and 4m	
		Return frequency between 4 and 6m	
		Return frequency between 6 and 8m	
		Return frequency between 8 and 10m	
		Return frequency above 10m	
Horizontal	Canopy cover		%
	Canopy density		%
Tree-centric	Number of trees		/
	Mean crown area		m^2
	Total crown area		m^2

2.4 Principal Component Analysis

First and foremost, the correlation matrix between variables was used to remove metrics that did not provide any additional information. Then, the eigenvalues of the Principal Component Analysis (PCA) were used to support the choice of the number of components to retain by keeping only the axis whose eigenvalue is greater than the mean (Palm, 2009). Finally, the correlation matrix of the variables with the selected axis, coupled with a significance test on each correlation value, allowed to choose the most significant variables per dimension.

In order to evaluate the usefulness of having many variables, three number of variables as input were tested: using all variables, selecting one variable per dimension or selecting two. The three approaches are discussed in the following sections.

2.5 Classification models

Three different classification models were developed: one clustering model and two discriminant models with different complexity levels. Before being used in classification, data were centered and reduced in order to standardize their weighting in the models.

The simplest model tested is a hierarchical clustering. The principle of this classification is to group most similar individuals then newly formed clusters together. Euclidean distance is the measure of distance used to characterize the dissimilarity between two individuals in the space of variables. The Ward's method was used to merge groups. Ward's criterion minimizes the total within-cluster variance.

For the two discriminant models, data were split into a training set and a testing set.

The next model built is a decision tree using the CART method. It works by recursively splitting the observations into groups of increasing homogeneity regarding their populations' distribution. The final result can be seen as a dichotomous key. The initial point is made out of the unclassified dataset and is called the root. The nodes are where the classification rule is applied. In this study, we have an univariate decision tree which means that each decision node consider the value of only one variable to create a split. The final groups are referred to as leaves and correspond to the four ecosystems.

The last model established is a Linear Discriminant Analysis (LDA). The assumptions of normality and homogeneity of within group covariance matrix were verified. The objective of this model is to create enough discriminant linear functions to delimit four regions, each representing one ecosystem. The discriminant linear functions are linear combinations of the previously selected variables. They are optimized to maximize the power of separation between groups. In this case, six discriminant linear functions were created, they represent the equations of six hyperplanes separating out the four ecosystems.

2.6 Upscaling

The objective of upscaling is to evaluate if the created models work on larger plots. As a reminder, the plots used to create the models are squares of 25m side. However, TERN works on 100m squares. The latter were divided into sub-sites of 25m side in order to increase the number of data. It is therefore interesting to determine if the built models are robust enough to work with plots of different sizes.

Only the two discriminant models were tested with plots of 100m side and plots of 50m side. The 100m squares are the 12 initial sites. Those same plots were cropped into 48 sub-sites to create the 50m squares. Thus, the test datasets contained 12 and 48 plots respectively. The accuracy of the confusion matrix for each model was used as a comparative parameter to assess the classification efficiency.

2.7 GEDI

The purpose of this comparison is to determine whether GEDI data (worldwide and open source) could be used as input to the models built in this project. To do this, two variables were extracted from GEDI data: a vertical metric, the maximum height, and a horizontal metric, the canopy cover. The descriptive statistics of those two metrics were then compared with the ones computed from the drone (M300) data in order to assess their similarities or dissimilarities.

The extraction of the GEDI data was done through Google Earth Engine. Raw GEDI data is a waveform representing the vertical profile of a 25m square. Due to the sparsity of the point cloud at the scale of the individual tree, GEDI data are not able to retrieve tree-centric parameter such as the number of trees (de Lame, 2021). On the Earth Engine Data Catalog, some processed GEDI metrics can be found, such as the maximum height and the canopy cover.

The GEDI satellite has been probing the Earth since 2018, but the entire globe is not yet covered. In order to find GEDI plots for the four ecosystems of this study, a buffer of 500m was drawn around each of the 12 initial sites. The assumption that the ecosystem remains the same 500m around each plot is therefore taken. This way, one GEDI plot was found for each site. The GEDI data extracted were collected between November 2020 and September 2022.

3 Results

In order to improve the readability of Figures and Tables, the four ecosystems will be referred to as chenopods (for chenopod shrubland), eucalypt (for eucalypt woodland), mallee and mulga.

3.1 Descriptive statistics on the variables

The descriptive statistics of the metrics can be observed on Figure 8. For each metric, ecosystems can be compared based on the median, the range and the data dispersion. Across all metrics, chenopod shrubland is the least variable ecosystem. On the other side, the ecosystem with the most variation is eucalypt woodland. Besides, chenopod shrubland stands out from the other ecosystems. On the one hand, for each metric, its median is lower. On the other hand, the ranges of the three other ecosystems systematically overlap.

Except for two parameters (number of trees and return frequency between 2 and 4m), the range of eucalypt woodland is wider than mallee and mulga in addition to covering most of their ranges. Mallee and mulga share quite similar characteristics. Indeed, except for the number of trees, there is a great overlap of their ranges. Another observation can be drawn: medians value of mallee and mulga are of the same order of magnitude except for two parameters (number of trees and total crown area).

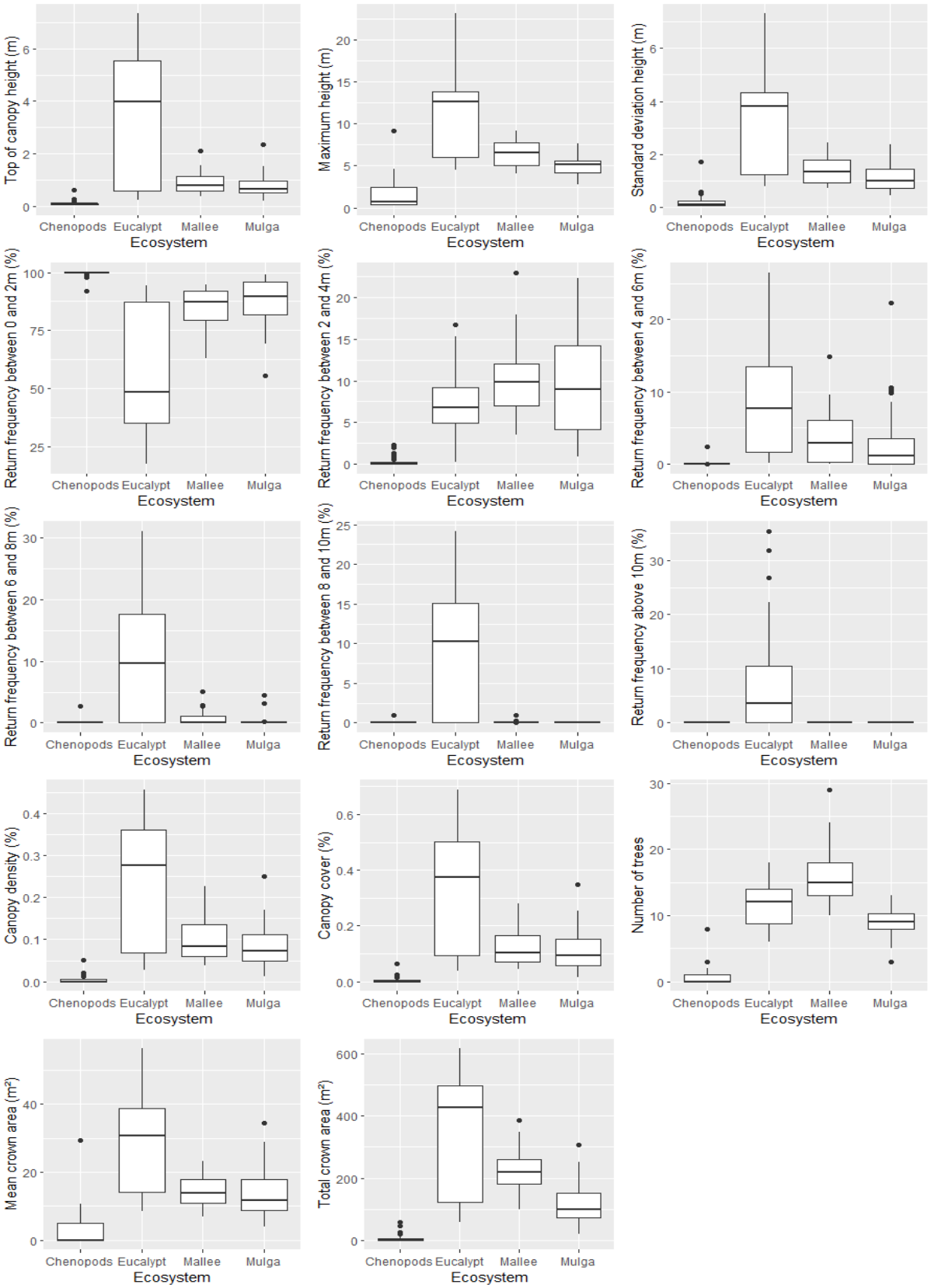

Figure 8: Descriptive statistics on the variables. For each metric, a comparison of the four ecosystems is made on the basis of boxplots. Boxplots include the data median, range and dispersion as well as outliers. The x axis represents the ecosystems: chenopods, eucalypt, mallee and mulga. The y axis represents each variable: top of canopy height, maximum height, standard deviation height, return frequency between 0 and 2m, return frequency between 2 and 4m, return frequency between 4 and 6m, return frequency between 6 and 8m, return frequency between 8 and 10m, return frequency above 10m, canopy density, canopy cover, number of trees, mean crown area and total crown area.

3.2 Principal Component Analysis

As previously stated in section 2.4, only the axis whose eigenvalue is greater than the mean were kept. Following this rule, the first two dimensions were retained. They explain 88% of the total cumulative variance of the plots. The most relevant metrics selected are: the top of canopy height and the density cover for the first dimension. And, the number of trees and the return frequency between 2 and 4m for the second dimension.

3.3 Classification models

3.3.1 Comparison of models with different numbers of variables

The developed classification models were a hierarchical clustering model, a decision tree and a LDA with a variable number of input metrics. For each model, three situations were implemented:

- all metrics as input;
- the top of canopy height, the canopy density, the number of trees and the return frequency between 2 and 4m as input;
- only the top of canopy height and the number of trees as input.

Accuracy related to the confusion matrix is used as a criterion for comparing models. Accuracies of models are summarized in Table 2: almost all models have an accuracy value above 75%. Based on this criterion, the worst model is the hierarchical clustering model with all input metrics (69% accuracy) whereas the best is the LDA with all input metrics (97% accuracy).

For hierarchical clustering, the more variables there are, the less accurate the model is. The opposite trend seems to apply to LDA whereas there is no clear pattern for decision trees.

Regardless of the number of input variables, hierarchical clustering is the model with the lowest accuracy and LDA has the highest accuracy. Moreover, the more input variables there are, the greater the accuracy difference between these two models.

Table 2: Comparison of models with different numbers of variables. Accuracies of the confusion matrices of the three models (hierarchical clustering, decision trees and LDA) with 14, 4 or 2 variables as input. The best model seems to be the LDA with 14 input variables.

	Hierarchical clustering	Decision trees	Linear Discriminant Analysis
14 variables (all)	0.69	0.88	0.97
4 variables	0.73	0.80	0.84
2 variables	0.80	0.83	0.84

3.3.2 Two variables models

Hierarchical clustering

Figure 9 shows the representation of each sub-plot in the first factorial plane of the PCA. Each sub-plot was classified in an ecosystem based on the hierarchical clustering model. The purple dots belong to the eucalypt woodland and are clearly separated from the other groups. The whole group is spread out as it occupies a large area on the plane. On the contrary, the blue dots, belonging to the chenopod shrubland, are all in the same place. The green dots belong to the mallee ecosystem and are also spread out but along the same line. Some overlapping can be noticed between the mallee and the mulga (in red).

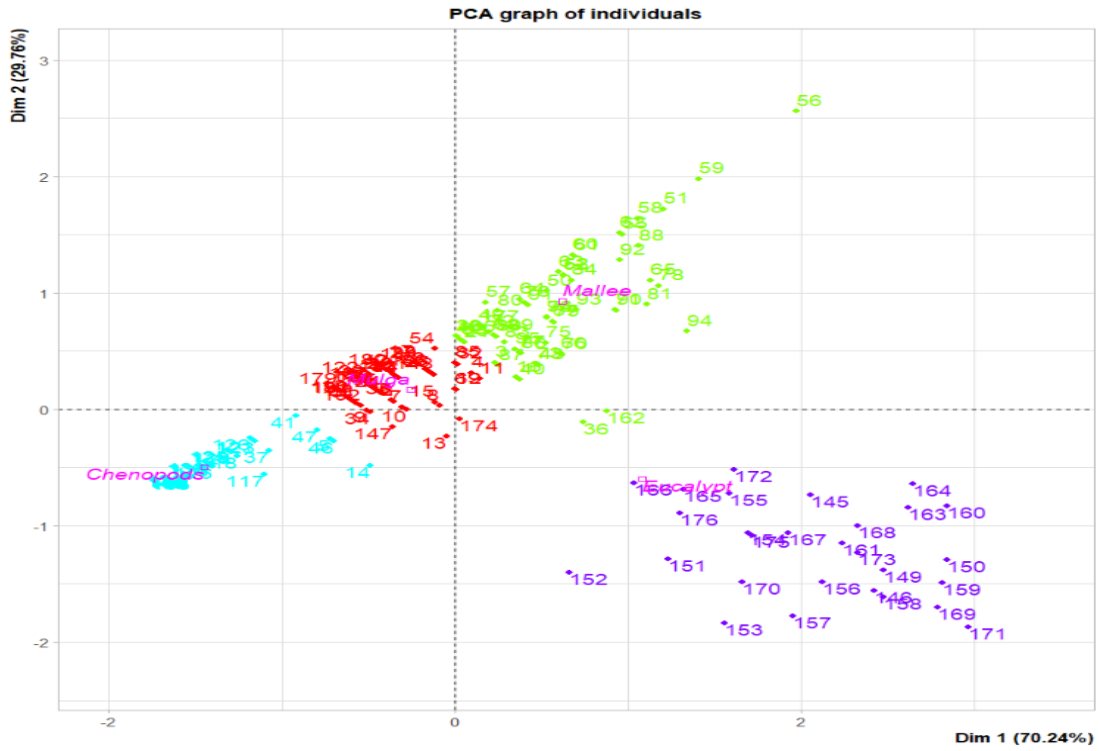

Figure 9: Plots classification using hierarchical clustering represented in the first factorial plane of the PCA. The x and y axis are the first and second dimension of the PCA respectively. Each dot represents a sub-plot. Each color is associated with an ecosystem: chenopods is blue, mulga is red, mallee is green and eucalypt is purple.

Table 3: Performance criteria of the hierarchical clustering. Sensitivities and specificities associated with each ecosystem (mulga, mallee, chenopods and eucalypt) as well as the accuracy of the confusion matrix for the hierarchical clustering classification. The overall accuracy is 80%.

	Mulga	Mallee	chenopods	Eucalypt
Sensitivity	0.69	0.94	0.98	0.58
Specificity	0.86	0.91	0.96	1.00
Accuracy	0.80			

Decision tree

Figure 10 represents the decision tree built from two variables, the top of canopy height and the number of trees, and applied on data validation. The first metric distinguishes individuals belonging to chenopod shrubland: any top of canopy height value below 0.3m is associated with this ecosystem. Classification of other ecosystems requires both metrics. On the one hand, the number of trees differentiates mallee from mulga. On the other hand, the top of canopy height is used to discriminate eucalypt woodland.

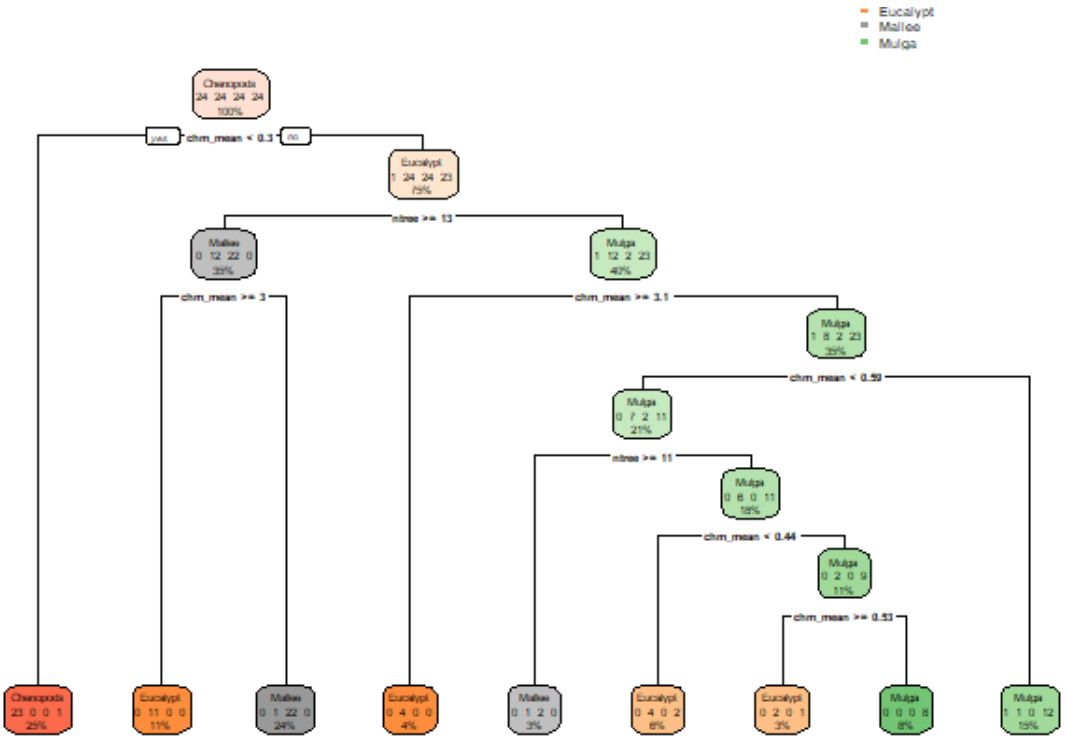

Figure 10: Decision tree classification. Dichotomous key applying thresholds on the two variables, the top of canopy height (chm_mean) and the number of trees (ntree), to create groups. Each color represents an ecosystem: chenopods are red, eucalypt is orange, mallee is grey and mulga is green.

Table 4: Performance criteria of the decision tree classification. Sensitivities and specificities associated with each ecosystem (mulga, mallee, chenopods and eucalypt) as well as the accuracy of the confusion matrix for the decision tree classification. The overall accuracy is 83%.

	Mulga	Mallee	chenopods	Eucalypt
Sensitivity	0.72	0.81	0.96	0.85
Specificity	0.92	0.96	1.00	0.91
Accuracy	0.83			

Linear Discriminant Analysis

Table 5: Performance criteria of the LDA classification. Sensitivities and specificities associated with each ecosystem (mulga, mallee, chenopods and eucalypt) as well as the accuracy of the confusion matrix for the LDA classification. The overall accuracy is 84%.

	Mulga	Mallee	chenopods	Eucalypt
Sensitivity	0.65	0.91	0.96	1.00
Specificity	0.97	0.96	1.00	0.88
Accuracy	0.84			

Tables 3, 4 and 5 describe the performance criteria for each model: the sensitivities and the specificities of each ecosystem as well as the global accuracy related to the confusion matrix. Sensitivity is the percentage of true positives whereas specificity is the percentage of true negatives. Overall, classification model accuracies are between 80% and 84%.

The following observations can be made regarding sensitivity rates:

– the value for mulga is less than 75% for all models;

– in each case, chenopod shrubland has the highest sensitivity, with a value close to 100% ($\geq$ 96%);

– for eucalypt woodland, hierarchical clustering leads to a low sensitivity (58%), while the discriminant models results in better values ($\geq$ 85%).

All specificity rates are above 85% which means that all three models accurately identify true negative. They are even up to 100% for eucalypt woodland and chenopod shrubland with hierarchical clustering and discriminant models respectively.

3.4 Upscaling

Classification models were created from square plots of 25m side. Subsequently, discriminant models based on two variables were used on larger plots (50m side and 100m side). In order to assess the classification efficiency, accuracy of the confusion matrix is used as a comparative parameter. Accuracies of models are summarized in Table 6. There is no significant difference in accuracy between the predictions on 50m squares or on 100m squares. The decision tree has an accuracy around 66% and the LDA, around 43%.

Table 6: Comparison of models accuracy tested with various plot sizes. Accuracies of the discriminant models (decision tree and LDA) trained using 25m square plots and tested with 50m or 100m square plots.

	Decision tree	Linear discriminant analysis
25m	0.83	0.84
50m	0.65	0.44
100m	0.67	0.42

3.5 GEDI

Figures 11 and 12 show the comparison between metrics (maximum height and canopy cover respectively) derived from drone (M300) data and from GEDI data. The same observations can be made for both metrics.

The data show statistical differences in terms of both dispersion and position parameters. Although there is partial overlap between the data, the following trends emerge except for chenopods: the GEDI data are less dispersed (lower standard deviation and range) and have a lower median. Furthermore, the GEDI data median is almost constant from one ecosystem to another (around 3.5m for maximum height and 0.03 for canopy cover).

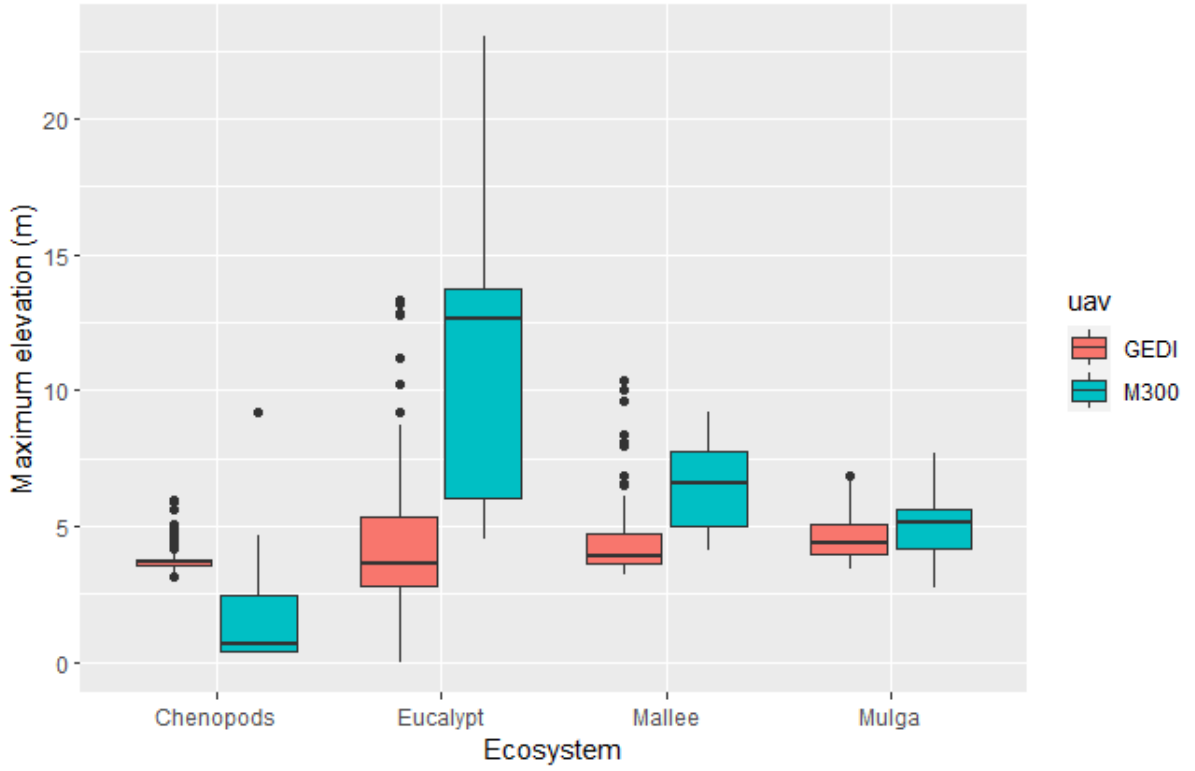

Figure 11: Comparison between the maximum height calculated from drone (M300) data and from GEDI data. Boxplots include the data median, range and dispersion as well as outliers. The x axis represents the ecosystems: chenopods, eucalypt, mallee and mulga. The y axis is the maximum elevation (m). The red boxplots are associated with GEDI and the blue ones with the drone (M300).

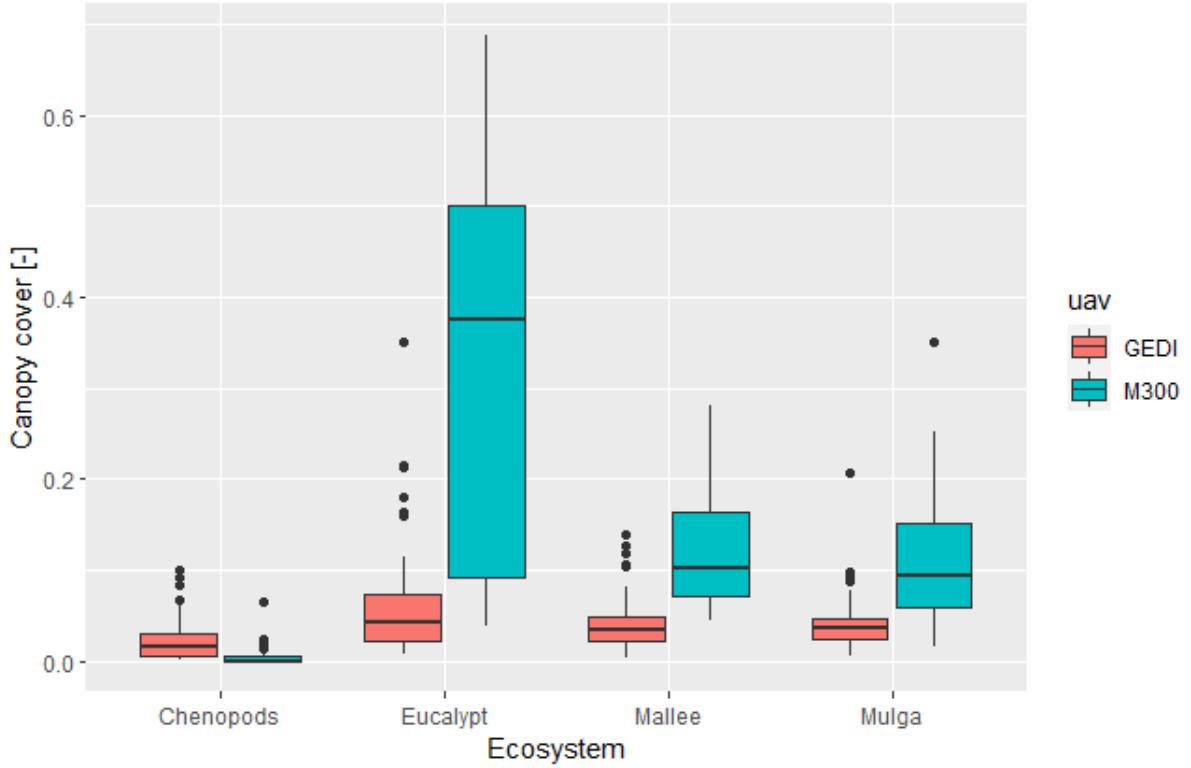

Figure 12: Comparison between the canopy cover calculated from drone (M300) data and from GEDI data. Boxplots include the data median, range and dispersion as well as outliers. The x axis represents the ecosystems: chenopods, eucalypt, mallee and mulga. The y axis is the canopy cover (-). The red boxplots are associated with GEDI and the blue ones with the drone (M300).

4 Discussion

4.1 Variables analysis

According to descriptive statistics, chenopod shrubland appears to be the most distinctive ecosystem and thus the easiest to differentiate. In the reality of the field, chenopod shrublands are indeed very different from the other three ecosystems since it has few to no tree. Even if a tree is found, it is usually a small shrub. The sparsity or absence of tree explain why the median of all metrics is lower for chenopod shrubland.

On the contrary, eucalypt woodland is the most variable ecosystem. It can therefore easily be misclassified. Indeed, its classification criteria in the literature ("NVIS Fact sheet MVG 5 – Eucalypt woodlands", 2017) are broader. Mallee and mulga are very specific and well-defined ecosystems while eucalypt woodland includes enormous range in composition and structure. The large number of eucalyptus species is already an example of its diversity.

For mallee and mulga, their descriptive characteristics are quite similar. This is therefore a possible source of misclassification. It would be interesting to do a significant difference test (Student's test) between the medians of each mallee and mulga variable to determine which parameter best discriminate the two ecosystems.

By comparing the LiDAR-derived maximum height with heights found in literature, eucalypt woodland, mallee and mulga are within their defined ranges, up to 30m, 10m and 10m respectively ("NVIS Fact sheet MVG 32 – Mallee open woodlands and sparse mallee shrublands", 2017; "NVIS Fact sheet MVG 5 – Eucalypt woodlands", 2017). However, some chenopod shrubland plots almost reach 5m and there is an outlier at about 9m while the NVIS Fact sheet MVG 22 fixes the maximum height at 2m. This might be explained by the presence of emergent trees which are sometimes found in chenopod shrubland.

From the PCA, it emerged that the variable that best characterizes the structure of ecosystems is the top of canopy height, which is consistent with literature (Asner et al., 2012; Campbell et al., 2017; Meyer et al., 2013).

4.2 Classification models

4.2.1 Comparison of models with different numbers of variables

The purpose of Table 2 is to evaluate the impact of the number of input metrics on models accuracy. The first observation that can be made is that two-variable models are more (or just as) accurate as four-variable models. This might be because the two added variables, the canopy density and the return frequency between 2 and 4m, provided more instability than additional information.

From Table 2, it can be concluded that hierarchical clustering is less effective with all variables than with only two metrics, whereas the opposite trend is observed for discriminant models. One hypothesis to explain the decrease in accuracy of hierarchical clustering with the number of metrics is that the more variables are taken into account, the more characteristics there are to differentiate one individual from another, but also the more sources of error. The model should be tested with a larger number of plots to verify its robustness. Of course, this applies to all models created in this study.

The difference in accuracy between the fourteen- and two-variable decision trees is only 5%. It is true that for decision trees, using a lot of input metrics brings instability. This is certainly why the all-variable model is not significantly more accurate than the two-variable model. To avoid this instability, we could switch to a random forest. Incidentally, Koma et al. (2021) demonstrated that a random forest is capable to differentiate land cover types and habitats within wetlands at high resolution using LiDAR-derived metrics from country-wide airborne laser scanning data. Reese et al. (2014) also successfully classified alpine vegetation using a random forest.

The fourteen-variable LDA is the most effective model for differentiating ecosystems, with an accuracy of 97%. The question of overfitting arises. Indeed, the more variables there are, the more fitted for that one dataset the discriminant equations are. Thus, a more parsimonious model can be expected to be more robust. Besides, the two-variable LDA has an accuracy of 84%, which is very good.

In itself, all two-variable models have an accuracy above 80% which is noteworthy. This means that simple models (with only two input parameters) can get fairly accurate classification results. It is often preferable to use a parsimonious model rather than a complex one because the final classification can be related to biological reality more easily (here: the link between the number of trees and their average height on a plot and the ecosystem which it belongs to). Therefore, it was decided to only evaluate the two-variable models to conduct the subsequent work. Their performance will be discussed in the next section.

4.2.2 Two variables models

Figure 9 shows a clear distinction between eucalypt woodland and the three other ecosystems. It is also seen that this group is quite spread out which points out a high within-cluster variance. This means that it is a fairly diversified ecosystem whose structural characteristics varies over a large range. This could already be observed through its descriptive statistics. On the contrary, the fact that chenopod shrubland is clustered at the same location indicates a low within-cluster variance. This is also confirmed by the small ranges of its descriptive statistics. As already said above, the sparsity or absence of trees in shrublands explains why this group does not have a lot of diversity and differs from the others. The overlapping between mallee and mulga noticed in Figure 9 could also be anticipated from their descriptive statistics. It stems from the fact that the two ecosystems share many similar characteristics. A final observation is that there is no site effect. Indeed, the plots come from two distant locations, but this distinction is not found in Figure 9, except slightly with eucalypt woodland. However, this separation will be discussed in the next paragraph.

From Tables 3, 4 and 5, the performance criteria of each model can be compared. On the one hand, sensitivity rates indicate that eucalypt woodland and mulga are the most difficult ecosystems to classify. Mulga is the hardest group to differentiate while the classification of eucalypt woodland strongly improves with discriminant models. Indeed, on the one side, its sensitivity for hierarchical clustering warns that there is a high misclassification rate. The visible separation on the factorial plane (Figure 9) is therefore erroneous. Many of the plots in this area are actually part of another ecosystem. Since eucalypt woodland metrics have the largest ranges, it is understandable that a few individuals were mistakenly assigned to this group. On the other side, its sensitivity for LDA is 100% which means that all individuals identified as eucalypt woodlands are indeed eucalypt woodlands. Also, its specificity informs that just a few eucalyptus plots were misclassified in another ecosystem. On the other hand, chenopod shrubland

is easily discriminated in all models. The decision tree (see Figure 10) corroborates this conclusion as it only applies one classification criterion to identify it.

All three model have an accuracy above 80%, so they can all efficiently differentiate the four ecosystems. Nonetheless, according to the performance criteria of each model, the LDA is the best suited model to discriminate these four ecosystems. The discriminating power of LiDAR-derived metrics coupled with an LDA had already been confirmed by Z. Zhang et al. (2011) who classified Australian temperate forests with an accuracy of 91%. Unfortunately, LDA is also the most complex model to interpret as the classification rules are made out of mathematical equations. On the contrary, the decision tree classification, which is only 1% less accurate than the LDA, is very easy to relate to the biological reality. Indeed, in Figure 10, we can clearly see which metric is used, between the top of canopy height and the number of trees, and which threshold is applied to identify a group.

4.2.3 Limitations

Even thought this study shows promising results, it is important to highlight its limitations.

One of the most important setback is the limited size of the dataset. Only 12 plots of 100m side were available and they had to be divided into smaller sub-sites to increase their statistical relevance. Only three sites represent each ecosystem which is not enough to capture their intraspecific diversity (or absence of it). It would be useful to test those models on a larger dataset in order to assess their robustness and identify a possible overfitting. Moreover, a larger dataset, and therefore a better representation of the ecosystems, might change the conclusions regarding which metrics best discriminate them.

I would also mention the fact that the plots used to train the models are a bit small. A square of 25m of side, 625m² of surface, is not ideal to represent an ecosystem. Small local variations will have a large effect on the plot characteristics. For example, if a small area does not contain a tree, it will be classified as chenopod shrubland. This scenario is quite common in open canopy forests and it might explain some of the misclassifications. It is therefore preferable to have large plots to prevent local variations from impacting the classification. It would be interesting to compare the descriptive statistics of the 100m squares to see how they differ from those of the 25m squares. But, once again, there are only 12 plots which is not enough to fairly represent an ecosystem.

Another limitation is the lack of field measurements to evaluate the quality of the metrics calculations. Unfortunately, I had no way (except the literature) to verify the accuracy of the parameters computed from LiDAR data. Calibration of LiDAR-derived variables with field data, such as the maximum height or the number of trees, would certainly improves the models. Indeed, remote sensing proxies will often need to be combined with field measurements to accurately represent the desired ecosystem (Tong et al., 2004).

Still in order to have LiDAR metrics that best reflect the reality, the calibration of the computing of the number of trees could be improved. As a reminder, the calculation of the number of trees requires the detection of local maxima thanks to a measurement window. In order to avoid the most omission and commission errors, it is important that the size of the measurement window is proportional to the area of the tree crown (Popescu & Wynne, 2004). Knowing that the crown area is directly proportional to the size of the tree, using a moving measurement window which varies with the height is appropriate. The equation (1) developed by Blanchard et al. (2016) was tested but unsuccessful. This might be due to the fact that it was created for tropical forests. Consequently, it does not perform as well on open canopy forests in drylands. However, the approach of using a moving window size seems pertinent. A different equation found in the literature could be used to increase the accuracy of the number of trees, which has proved to be an essential variable for discriminating ecosystems.

Also, I think that the great variability of eucalypt woodland highly influences the accuracy of the models. Indeed, as this ecosystem is very diversified, many plots can be found there. To compensate for this deficiency, we could divide it into narrower and more precisely defined ecosystems.

Lastly, the only criterion to evaluate model performance was the confusion matrix, but it would be appropriate to quantify their uncertainty with other parameters as well.

4.3 Upscaling

The upscaling was carried out to determine if the models work on larger plots, especially on 100m squares as this is the size on which TERN works. The original 12 plots had to be cropped into multiple sub-sites in order to increase the size of the dataset. Based on the results presented in Table 6, it can be concluded that models are less effective on larger plots (regardless of size). The LDA lost half its accuracy while the decision tree dropped to 66% accuracy. It appears that this model is more robust but still not effective enough to properly differentiate the four ecosystems in this study.

In fact, it makes perfect sense that the models are inaccurate since one of the input variables, the number of trees, depends on the size of the plot. It is obvious that a 10 000m² plot and a 625m² plot do not contain the same number of trees. Only the top of canopy height can be transposed to larger scale models. Thus, the models built in this study can only be applied to 25m squares. The ideal would have been to create the models from the 100m squares but only 12 data is not enough to assess the model quality. Perhaps using another metric, which is not dependent on the plot size, would improve the models. The return frequency between 2 and 4m, which was the second variable selected for the second dimension of the PCA, could be a good lead.

4.4 GEDI

The comparison between the drone metrics and the GEDI metrics (Figures 11 and 12) shows that the GEDI metrics are less variable than the drone metrics. Besides, the GEDI medians are constant across the four ecosystems. These conclusions apply to both the vertical variable, maximum height, and to the horizontal variable, canopy cover. This means that GEDI metrics are unfit to discriminate these ecosystems with the method established in this study.

One hypothesis to explain the observed differences is that the GEDI data come from a satellite: its resolution and, therefore, its accuracy are much lower than those of the drone. A satellite is less able to detect small variations and therefore the ranges of GEDI variables are smaller. Another explanation for the contrast between the two datasets is that the GEDI data were collected under the assumption that the ecosystem remains the same 500m around each site. Presumably it might not have been the case for some plots, which could explain the dissimilarity between metrics. What is more, the GEDI variables were found on the Earth Engine Data Catalog; their method of calculation may differ from the way I computed them.

4.5 Additional prospects

The use of LiDAR for large-scale ecosystems characterization is promising. LiDAR provides detailed information on the three-dimensional structure of vegetation. Unlike some other technologies such as optical satellite imagery, LiDAR can penetrate through vegetation to measure the surface area of the ground, allowing a better understanding of the underlying topography. LiDAR can distinguish different vegetation layers and even detect individual trees. But, it also has disadvantages. LiDAR technology can be costly to implement in terms of data collection, processing and analysis. The equipment required is often sophisticated. Weather conditions, such as rain or fog, may affect the quality of the data. In addition, raw LiDAR data can be complex to interpret and require advanced technical skills to be properly processed and analyzed.

This study is an example proving that LiDAR technology can be used to discriminate ecosystems but it is not the only one. Indeed, LiDAR-derived metrics can successfully classify forest structural classes (Campbell et al., 2017; Jones et al., 2012) or wetland-related land cover types and habitats (Koma et al., 2021). In Australia, Z. Zhang et al. (2011) accurately classified temperate forests. Even in area with high uncertainty such as semi-arid lands, small-footprint waveform features can characterize heterogeneous vegetation at high spatial resolution. Pulse width appeared to be a relevant LiDAR metric as it allowed to differentiate bare ground from low-height vegetation (Ilangakoon et al., 2018).

However, combining different remote sensing sources seems to enhance the results. Indeed, Trouvé et al. (2023) found that combining environmental, multispectral and LiDAR data improves the classification of temperate forest in Australia. What is more, the fusion of airborne laser scanning data and optical satellite data gave the highest classification accuracy in alpine region (Reese et al., 2014). Dashti et al. (2019) employed spectral and structural (LiDAR) data to circumvent the challenges associated with drylands. Incorporating the LiDAR information into the classification scheme increased the overall accuracy from 60% to 89%.

5 Conclusion

This study demonstrated the potential use of high resolution airborne LiDAR data to classify four ecosystems found in the outback of South Australia. The characterization of ecosystems is done by describing vegetation structure. Based on numerous studies on the subject (Dubayah & Drake, 2000; Huylenbroeck et al., 2020; Lim et al., 2003; Michez et al., 2016), we hypothesized that the structural components of vegetation could be determined using LiDAR technology.

Various LiDAR-derived metrics were computed but a PCA help select the best suited ones in order to differentiate our ecosystems. Three classification models were created: a hierarchical clustering, a decision trees and a LDA. A comparison between the models built with different numbers of variables as input revealed that only two variables, the top of canopy height and the number of trees, were enough to accurately identify the ecosystems. Indeed, all three of them have an accuracy equal or above 80%. A more detailed analysis of the models performance indicated that eucalypt woodland and mulga are the hardest ecosystems to classify while chenopod shrubland is the easiest. The LDA appeared to be the best predictor but the decision tree, whose overall accuracy is only 1% less, is easier to interpret and to relate to ecological reality.

Trying out the models with plots of larger sizes was not concluding because the number of trees, one of the input variable, is dependent on the plot size. Therefore, the models created in this study can only be used with 25m squares.

Another way of upscaling the scope of this study was explored through the use of GEDI data, which are collected worldwide on 25m squares. In theory, applying GEDI data as input would allow for a large-scale use of our models. Unfortunately, the comparison of airborne and spaceborne LiDAR metrics revealed a significant difference between the two datasets.

Even though this work showed promising results in using LiDAR technology to characterize ecosystems, there is always room for improvement. Testing the models on a larger dataset would assess their robustness and determine if they can be transposed to other study areas. What is more, a calibration of the LiDAR-derived metrics with field measurements would enhance the quality of the metrics. The number of trees calculation also revealed a need for better fitted methods in drylands. In addition, combining different sources of remote sensing proved to increase classification accuracy.

This study showed great results in using high resolution airborne LiDAR to discriminate some Australian ecosystems. Nonetheless, the limited amount of data calls for further research in order to improve the large-scale characterization of drylands ecosystems.

6 Contribution of the student

In terms of field data collection, I only participated in the Wintinna campaign. The drone flights were conducted by Nick Gellie from the TERN team. I assisted him in the preparation of the flights as well as during the data acquisition where I monitored the drone's safety by verifying the wind speed as well as a possible bird attack.

I processed the data from the raw drone images to the derived-LiDAR metrics. However, I would like to point out that an R code was provided to me by Poornima Sivanandam, from the University of Tasmania, who works in collaboration with TERN, to create CHMs from .las files. This code also contained the calculation of canopy cover and canopy density. I adapted it to the needs of this study. All other variables, statistical analysis and classification models are purely the object of my creation. I also did the extraction and process of the GEDI data.

In addition, the data analysis, the interpretation of the results, the writing of this document and the communication of the results were done by myself.

7 References

Adeel, Z. (2005). *Ecosystems and human well-being: Desertification synthesis: A report of the millennium ecosystem assessment* (W. R. Institute, Ed.). World Resources Institute.

Asner, G. P., Clark, J. K., Mascaro, J., Galindo García, G. A., Chadwick, K. D., Navarrete Encinales, D. A., Paez-Acosta, G., Cabrera Montenegro, E., Kennedy-Bowdoin, T., Duque, Á., Balaji, A., von Hildebrand, P., Maatoug, L., Phillips Bernal, J. F., Yepes Quintero, A. P., Knapp, D. E., García Dávila, M. C., Jacobson, J., & Ordóñez, M. F. (2012). High-resolution mapping of forest carbon stocks in the colombian amazon [Publisher: Copernicus GmbH]. *Biogeosciences*, *9*(7), 2683–2696. https://doi.org/10.5194/bg-9-2683-2012

Asner, G. P., & Mascaro, J. (2014). Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sensing of Environment*, *140*, 614–624. https://doi.org/10.1016/j.rse.2013.09.023

Bansod, B., Singh, R., Thakur, R., & Singhal, G. (2017). A comparision between satellite based and drone based remote sensing technology to achieve sustainable development: A review [Number: 2]. *Journal of Agriculture and Environment for International Development (JAEID)*, *111*(2), 383–407. https://doi.org/10.12895/jaeid.20172.690

Bastin, J.-F., Berrahmouni, N., Grainger, A., Maniatis, D., Mollicone, D., Moore, R., Patriarca, C., Picard, N., Sparrow, B., Abraham, E. M., Aloui, K., Atesoglu, A., Attore, F., Bassüllü, Ç., Bey, A., Garzuglia, M., García-Montero, L. G., Groot, N., Guerin, G., … Castro, R. (2017). The extent of forest in dryland biomes [Publisher: American Association for the Advancement of Science]. *Science*, *356*(6338), 635–638. https://doi.org/10.1126/science.aam6527

Blanchard, E., Birnbaum, P., Ibanez, T., Boutreux, T., Antin, C., Ploton, P., Vincent, G., Pouteau, R., Vandrot, H., Hequet, V., Barbier, N., Droissart, V., Sonké, B., Texier, N., Kamdem, N. G., Zebaze, D., Libalah, M., & Couteron, P. (2016). Contrasted allometries between stem diameter, crown area, and tree height in five tropical biogeographic areas. *Trees*, *30*(6), 1953–1968. https://doi.org/10.1007/s00468-016-1424-3

Cabello, J., Fernández, N., Alcaraz-Segura, D., Oyonarte, C., Piñeiro, G., Altesor, A., Delibes, M., & Paruelo, J. M. (2012). The ecosystem functioning dimension in conservation: Insights from remote sensing. *Biodiversity and Conservation*, *21*(13), 3287–3305. https://doi.org/10.1007/s10531-012-0370-7

Campbell, L., Coops, N. C., & Saunders, S. C. (2017). LiDAR as an advanced remote sensing technology to augment ecosystem classification and mapping [Number: 1]. *Journal of Ecosystems and Management*, *17*(1). https://doi.org/10.22230/jem.2017v17n1a588

Chave, J., Davies, S. J., Phillips, O. L., Lewis, S. L., Sist, P., Schepaschenko, D., Armston, J., Baker, T. R., Coomes, D., Disney, M., Duncanson, L., Hérault, B., Labrière, N., Meyer, V., Réjou-Méchain, M., Scipal, K., & Saatchi, S. (2019). Ground data are essential for biomass remote sensing missions. *Surveys in Geophysics*, *40*(4), 863–880. https://doi.org/10.1007/s10712-019-09528-w

Coomes, D. A., Dalponte, M., Jucker, T., Asner, G. P., Banin, L. F., Burslem, D. F. R. P., Lewis, S. L., Nilus, R., Phillips, O. L., Phua, M.-H., & Qie, L. (2017). Area-based vs tree-centric approaches to mapping forest carbon in southeast asian forests from airborne laser scanning data. *Remote Sensing of Environment*, *194*, 77–88. https://doi.org/10.1016/j.rse.2017.03.017

Dalponte, M., & Coomes, D. A. (2016). Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12575]. *Methods in Ecology and Evolution*, *7*(10), 1236–1245. https://doi.org/10.1111/2041-210X.12575

Dashti, H., Poley, A., F. Glenn, N., Ilangakoon, N., Spaete, L., Roberts, D., Enterkine, J., N. Flores, A., L. Ustin, S., & J. Mitchell, J. (2019). Regional scale dryland vegetation classification with an integrated lidar-hyperspectral approach [Number: 18 Publisher: Multidisciplinary Digital Publishing Institute]. *Remote Sensing*, *11*(18), 2141. https://doi.org/10.3390/rs11182141

de Lame, H. (2021). Characterization of tropical forest structure using airborne LiDAR in central Africa.

Dubayah, R. O., & Drake, J. B. (2000). Lidar remote sensing for forestry [Num Pages: 3 Place: Bethesda, United Kingdom Publisher: Oxford University Press], *98*(6), 44. Retrieved August 15, 2023, from https://www.proquest.com/docview/220790008/citation/711C5D43E151409DPQ/1

Durant, S. M., Pettorelli, N., Bashir, S., Woodroffe, R., Wacher, T., De Ornellas, P., Ransom, C., Abáigar, T., Abdelgadir, M., El Alqamy, H., Beddiaf, M., Belbachir, F., Belbachir-Bazi, A., Berbash, A. A., Beudels-Jamar, R., Boitani, L., Breitenmoser, C., Cano, M., Chardonnet, P., … Baillie, J. E. M. (2012). Forgotten biodiversity in desert ecosystems. *Science*, *336*(6087), 1379–1380. https://doi.org/10.1126/science.336.6087.1379

Eamus, D., Huete, A., Cleverly, J., Nolan, R. H., Ma, X., Tarin, T., & Santini, N. S. (2016). Mulga, a major tropical dry open forest of australia: Recent insights to carbon and water fluxes [Publisher: IOP Publishing]. *Environmental Research Letters*, *11*(12), 125011. https://doi.org/10.1088/1748-9326/11/12/125011

Ferraz, A., Saatchi, S., Mallet, C., & Meyer, V. (2016). Lidar detection of individual tree size in tropical forests. *Remote Sensing of Environment*, *183*, 318–333. https://doi.org/10.1016/j.rse.2016.05.028

Foulkes, J., de Preu, N., Sinclair, R., Thurgate, N., Sparrow, B., & White, A. (2014, January 1). Chenopod and acacia shrublands.

Ganem, K. A., Xue, Y., Rodrigues, A. d. A., Franca-Rocha, W., Oliveira, M. T. d., Carvalho, N. S. d., Cayo, E. Y. T., Rosa, M. R., Dutra, A. C., & Shimabukuro, Y. E. (2022). Mapping south america's drylands through remote sensing—a review of the methodological trends and current challenges [Number: 3 Publisher: Multidisciplinary Digital Publishing Institute]. *Remote Sensing*, *14*(3), 736. https://doi.org/10.3390/rs14030736

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., & Sohlberg, R. A. (2003). Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS veg-

etation continuous fields algorithm [Publisher: American Meteorological Society Section: Earth Interactions]. *Earth Interactions*, *7*(10), 1–15. https://doi.org/10.1175/1087-3562(2003)007<0001:GPTCAA>2.0.CO;2

Höfle, B., & Hollaus, M. (2010). Urban vegetation detection using high density full-waveform airborne LiDAR data - combination of object-based image and point cloud analysis. *38*, 281–286. Retrieved July 13, 2023, from https://publik.tuwien.ac.at/files/PubDat_188128.pdf

Huang, J., Yu, H., Guan, X., Wang, G., & Guo, R. (2016). Accelerated dryland expansion under climate change [Num Pages: 7 Place: London, United States Publisher: Nature Publishing Group], *6*(2), 166–171. https://doi.org/10.1038/nclimate2837

Huylenbroeck, L., Laslier, M., Dufour, S., Georges, B., Lejeune, P., & Michez, A. (2020). Using remote sensing to characterize riparian vegetation: A review of available tools and perspectives for managers. *Journal of Environmental Management*, *267*, 110652. https://doi.org/10.1016/j.jenvman.2020.110652

Hyyppa, J., Kelle, O., Lehikoinen, M., & Inkinen, M. (2001). A segmentation-based method to retrieve stem volume estimates from 3-d tree height models produced by laser scanners [Conference Name: IEEE Transactions on Geoscience and Remote Sensing]. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(5), 969–975. https://doi.org/10.1109/36.921414

Ilangakoon, N. T., Glenn, N. F., Dashti, H., Painter, T. H., Mikesell, T. D., Spaete, L. P., Mitchell, J. J., & Shannon, K. (2018). Constraining plant functional types in a semi-arid ecosystem with waveform lidar. *Remote Sensing of Environment*, *209*, 497–509. https://doi.org/10.1016/j.rse.2018.02.070

Jones, T. G., Coops, N. C., & Sharma, T. (2012). Assessing the utility of LiDAR to differentiate among vegetation structural classes [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01431161.2011.559289
Remote Sensing Letters, *3*(3), 231–238. https://doi.org/10.1080/01431161.2011.559289

Keith, D. A. (2017, June 15). *Australian vegetation* [Google-Books-ID: Dk3ODgAAQBAJ]. Cambridge University Press.

Kellner, J. R., Armston, J., Birrer, M., Cushman, K. C., Duncanson, L., Eck, C., Falleger, C., Imbach, B., Král, K., Krůček, M., Trochta, J., Vrška, T., & Zgraggen, C. (2019). New opportunities for forest remote sensing through ultra-high-density drone lidar. *Surveys in Geophysics*, *40*(4), 959–977. https://doi.org/10.1007/s10712-019-09529-9

Ko, D., Bristow, N., Greenwood, D., & Weisberg, P. (2009). Canopy cover estimation in semiarid woodlands: Comparison of field-based and remote sensing methods. *Forest Science*, *55*(2), 132–141.

Koma, Z., Seijmonsbergen, A. C., & Kissling, W. D. (2021). Classifying wetland-related land cover types and habitats using fine-scale lidar metrics derived from country-wide airborne laser scanning [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rse2.170]. *Remote Sensing in Ecology and Conservation*, *7*(1), 80–96. https://doi.org/10.1002/rse2.170

Leitold, V., Keller, M., Morton, D. C., Cook, B. D., & Shimabukuro, Y. E. (2015). Airborne lidar-based estimates of tropical forest structure in complex terrain: Opportunities and trade-offs for REDD+. *Carbon Balance and Management*, *10*(1), 3. https://doi.org/10.1186/s13021-015-0013-x

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure [Publisher: Sage Publications, Ltd.]. *Progress in Physical Geography*, *27*(1), 88–106. https://doi.org/10.1191/0309133303pp360ra

Meyer, V., Saatchi, S. S., Chave, J., Dalling, J. W., Bohlman, S., Fricker, G. A., Robinson, C., Neumann, M., & Hubbell, S. (2013). Detecting tropical forest biomass dynamics from repeated airborne lidar measurements [Publisher: Copernicus GmbH]. *Biogeosciences*, *10*(8), 5421–5438. https://doi.org/10.5194/bg-10-5421-2013

Michez, A., Bauwens, S., Bonnet, S., & Lejeune, P. (2016). Characterization of forests with LiDAR technology. In *Land surface remote sensing in agriculture and forest* (pp. 331–362). Elsevier. https://doi.org/10.1016/B978-1-78548-103-1.50008-X

Miura, N., & Jones, S. D. (2010). Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. *Remote Sensing of Environment*, *114*(5), 1069–1076. https://doi.org/10.1016/j.rse.2009.12.017

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities [Number: 6772 Publisher: Nature Publishing Group]. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Nagendra, H., Lucas, R., Honrado, J. P., Jongman, R. H. G., Tarantino, C., Adamo, M., & Mairota, P. (2013). Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators*, *33*, 45–59. https://doi.org/10.1016/j.ecolind.2012.09.014

NVIS fact sheet MVG 16 – acacia shrublands. (2017). *Australian Government, Department of the Environment and Energy*.

NVIS fact sheet MVG 22 – chenopod shrublands, samphire shrublands and forblands. (2017). *Australian Government, Department of the Environment and Energy*.

NVIS fact sheet MVG 32 – mallee open woodlands and sparse mallee shrublands. (2017). *Australian Government, Department of the Environment and Energy*.

NVIS fact sheet MVG 5 – eucalypt woodlands. (2017). *Australian Government, Department of the Environment and Energy*.

Palm, R. (2009). *L'analyse en composante principales: Principes et applications.* Université de Liège – Gembloux Agro-Bio Tech.

Peart, D. R. (1990). Review of mediterranean landscapes in australia: Mallee ecosystems and their management. [Publisher: University of Chicago Press]. *The Quarterly Review of Biology*, *65*(3), 376–377. Retrieved July 10, 2023, from https://www.jstor.org/stable/2832433

Pettorelli, N., Laurance, W. F., O'Brien, T. G., Wegmann, M., Nagendra, H., & Turner, W. (2014). Satellite remote sensing for applied ecologists: Opportunities and challenges [_eprint: https://onlinelibrary.wiley.com/d 2664.12261]. *Journal of Applied Ecology*, *51*(4), 839–848. https://doi.org/10.1111/1365-2664.12261

Popescu, S. C., & Wynne, R. H. (2004). Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering & Remote Sensing*, *70*(5), 589–604.

Popescu, S. C., Wynne, R. H., & Nelson, R. F. (2002). Estimating plot-level tree heights with lidar: Local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture*, *37*(1), 71–95. https://doi.org/10.1016/S0168-1699(02)00121-7

Reese, H., Nyström, M., Nordkvist, K., & Olsson, H. (2014). Combining airborne laser scanning data and optical satellite data for classification of alpine vegetation. *International Journal of Applied Earth Observation and Geoinformation*, *27*, 81–90. https://doi.org/10.1016/j.jag.2013.05.003

Roussel, J.-R., Auty, D., Coops, N. C., Tompalski, P., Goodbody, T. R. H., Meador, A. S., Bourdon, J.-F., de Boissieu, F., & Achim, A. (2020). lidR: An r package for analysis of airborne laser scanning (ALS) data. *Remote Sensing of Environment*, *251*, 112061. https://doi.org/10.1016/j.rse.2020.112061

Ruusa, D. M., Rosser, N. J., & Donoghue, D. N. M. (2022). Remote sensing for monitoring tropical dryland forests: A review of current research, knowledge gaps and future directions for southern africa [Publisher: IOP Publishing]. *Environmental Research Communications*, *4*(4), 042001. https://doi.org/10.1088/2515-7620/ac5b84

Schneider, F. D., Ferraz, A., Hancock, S., Duncanson, L. I., Dubayah, R. O., Pavlick, R. P., & Schimel, D. S. (2020). Towards mapping the diversity of canopy structure from space with GEDI [Publisher: IOP Publishing]. *Environmental Research Letters*, *15*(11), 115006. https://doi.org/10.1088/1748-9326/ab9e99

Sexton, J. O., Noojipady, P., Song, X.-p., Feng, M., Song, D.-x., Kim, D.-h., Anand, A., Huang, C., Channan, S., Pimm, S. L., & Townshend, J. R. (2016). Conservation policy and the measurement of forests [Num Pages: 6 Place: London, United States Publisher: Nature Publishing Group], *6*(2), 192–196. https://doi.org/10.1038/nclimate2816

Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K. M., Channan, S., DiMiceli, C., & Townshend, J. R. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17538947.2013.786146]. *International Journal of Digital Earth*, *6*(5), 427–448. https://doi.org/10.1080/17538947.2013.786146

Silva, C. A., Hudak, A. T., Vierling, L. A., Loudermilk, E. L., O'Brien, J. J., Hiers, J. K., Jack, S. B., Gonzalez-Benecke, C., Lee, H., Falkowski, M. J., & Khosravipour, A. (2016). Imputation of individual longleaf pine (pinus palustris mill.) tree attributes from field and LiDAR data [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07038992.2016.1196582]. *Canadian Journal of Remote Sensing*, *42*(5), 554–573. https://doi.org/10.1080/07038992.2016.1196582

Smith, W. K., Dannenberg, M. P., Yan, D., Herrmann, S., Barnes, M. L., Barron-Gafford, G. A., Biederman, J. A., Ferrenberg, S., Fox, A. M., Hudson, A., Knowles, J. F., MacBean, N., Moore, D. J. P., Nagler, P. L., Reed, S. C., Rutherford, W. A., Scott, R. L., Wang, X., & Yang, J. (2019). Remote sensing of dryland ecosystem structure and function: Progress, challenges, and opportunities. *Remote Sensing of Environment*, *233*, 111401. https://doi.org/10.1016/j.rse.2019.111401

Sparrow, B. D., Edwards, W., Munroe, S. E., Wardle, G. M., Guerin, G. R., Bastin, J.-F., Morris, B., Christensen, R., Phinn, S., & Lowe, A. J. (2020). Effective ecosystem monitoring requires a multi-scaled approach [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/brv.12636]. *Biological Reviews*, *95*(6), 1706–1719. https://doi.org/10.1111/brv.12636

Tong, C., Wu, J., Yong, S., Yang, J., & Yong, W. (2004). A landscape-scale assessment of steppe degradation in the xilin river basin, inner mongolia, china. *Journal of Arid Environments*, *59*(1), 133–149. https://doi.org/10.1016/j.jaridenv.2004.01.004

Trouvé, R., Jiang, R., Fedrigo, M., White, M. D., Kasel, S., Baker, P. J., & Nitschke, C. R. (2023). Combining environmental, multispectral, and LiDAR data improves forest type classification: A case study on mapping cool temperate rainforests and mixed forests [Number: 1 Publisher: Multidisciplinary Digital Publishing Institute]. *Remote Sensing*, *15*(1), 60. https://doi.org/10.3390/rs15010060

Williams, R. J., Gill, A. M., & Bradstock, R. A. (2012, February 21). *Flammable australia: Fire regimes, biodiversity and ecosystems in a changing world* [Google-Books-ID: DghUhpiiVCsC]. Csiro Publishing.

Zhang, J., Guo, W., Zhou, B., & Okin, G. S. (2021). Drone-based remote sensing for research on wind erosion in drylands: Possible applications [Number: 2 Publisher: Multidisciplinary Digital Publishing Institute]. *Remote Sensing*, *13*(2), 283. https://doi.org/10.3390/rs13020283

Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X., & Yan, G. (2016). An easy-to-use airborne LiDAR data filtering method based on cloth simulation [Number: 6 Publisher: Multidisciplinary Digital Publishing Institute]. *Remote Sensing*, *8*(6), 501. https://doi.org/10.3390/rs8060501

Zhang, Z., Liu, X., Peterson, J., & Wright, W. (2011). Cool temperate rainforest and adjacent forests classification using airborne LiDAR data [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-4762.2011.01035.x]. *Area*, *43*(4), 438–448. https://doi.org/10.1111/j.1475-4762.2011.01035.x

8 Appendices

Table 7: Site name, latitude and longitude of the up right corner of the plot, location and ecosystem of each site.

Name	Latitude [°]	Longitude [°]	Location	Ecosystem
sasmdd0001	-34,001805	140,587086	Calperum	Mallee
sasmdd0002	-34,010801	140,593779	Calperum	Mallee
sasmdd0003	-33,993633	140,586725	Calperum	Mallee
sasmdd0005	-33,973407	140,726868	Calperum	Chenopods
sasmdd0008	-34,043302	140,761086	Calperum	Eucalypt
sasmdd0011	-34,017995	140,710609	Calperum	Chenopods
sasmdd0012	-34,058235	140,755431	Calperum	Eucalypt
sasmdd0013	-34,055805	140,735956	Calperum	Eucalypt
saagvd0005	-27,706331	133,673490	Wintinna	Mulga
saagvd0007	-27,668755	133,725231	Wintinna	Mulga
saagvd0008	-27,638147	133,767499	Wintinna	Mulga
saastp0036	-27,569344	134,144672	Wintinna	Chenopods

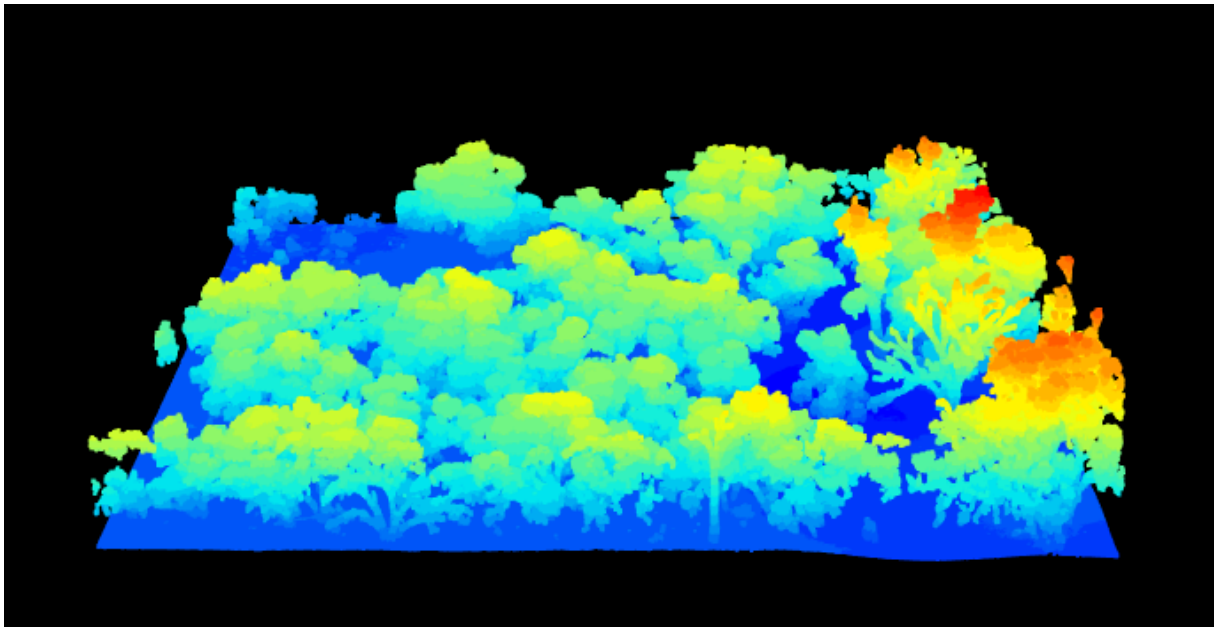

Figure 13: Point cloud of sasmdd0008 constructed using the DJI Terra software.