Using topographic attributes to predict the understorey structure of a wet eucalypt forest

Bechu K.V. Yadav\textsuperscript{a*}, Arko Lucieer\textsuperscript{a}, Gregory J. Jordan\textsuperscript{b}, Susan C. Baker\textsuperscript{b,c}

\textsuperscript{a} University of Tasmania, School of Technology, Environments and Design, Discipline of Geography and Spatial Sciences, Private Bag 70, Hobart, Tasmania 7001, Australia.

\textsuperscript{b} University of Tasmania, School of Natural Sciences, Biological Sciences, Private Bag 55, Hobart, Tasmania 7001, Australia.

\textsuperscript{c} ARC Centre for Forest Value, University of Tasmania, Hobart, Tasmania 7001, Australia.

*Corresponding author: University of Tasmania, School of Technology, Environments and Design, Discipline of Geography and Spatial Sciences, Private Bag 70, Hobart, Tasmania 7001, Australia. Tel: +61 3 6226 2140

E-mail addresses: bechu.yadav@utas.edu.au (B.K.V. Yadav), arko.lucieer@utas.edu.au (A. Lucieer), greg.jordan@utas.edu.au (G.J. Jordan), s.baker@utas.edu.au (S.C. Baker)
Abstract

**Background:** Forest understorey structure is an important component of forest ecosystems that affects forest-dwelling species, nutrient cycling, fire behaviour, biodiversity, and regeneration capacity. Mapping the structure of forest understorey vegetation with field surveys or high-resolution LiDAR data is costly. We tested whether landscape topography and underlying geology could predict the understorey structure of a 19 km² area of wet eucalypt primary forest located at the Warra Long Term Ecological Research Supersite, Tasmania, Australia. In this study, we used random forest regressions based on twelve topographic attributes derived from digital terrain models (DTMs) at various resolutions and a geology variable to predict the densities of three understorey layers compared to density estimates from a high resolution (28.66 points/m²) LiDAR survey.

**Results:** We predicted the vegetation density of three canopy strata with a high degree of accuracy (validation root mean square error ranged from 8.97% to 13.69%). 30 m resolution DTMs provided greater predictive accuracy than DTMs with higher spatial resolution. Variable importance depended on spatial resolutions and canopy strata layers, but among the predictor variables, geology generally produced the highest predictive importance followed by solar radiation. Topographic position index, aspect, and SAGA wetness index had moderate importance.

**Conclusions:** This study demonstrates that geological and topographic attributes can provide useful predictions of understorey vegetation structure in a primary forest. Given the good performance of 30 m resolution, the predictive power of the models could be tested on a larger geographical area using lower density.
LiDAR point clouds. This study should help in assessing fuel loads, carbon stores, biomass, and biological diversity, and could be useful for foresters and ecologists contributing to the planning of sustainable forest management and biodiversity conservation.

**Keywords:** Airborne LiDAR, Digital terrain model, Topographic attributes, Geology, Understorey structure, Random forest, Variable importance

**Introduction**

Forest understorey vegetation is an essential component of forest ecosystems (Latifi et al. 2017) that provides wildlife habitat and influences fire behaviour, nutrient cycling, biodiversity, and regeneration potential (Campbell et al. 2018; Simonson et al. 2014; Wing et al. 2012). In many temperate forest ecosystems, most of the plant biodiversity is contained within the understorey vegetation layers (Weisberg et al. 2003). Moreover, the vertical structure of forest understoreys is a vital attribute that affects habitat quality, including foraging and breeding resources, for many forest animals (Camprodon and Brotons 2006). Thus, characteristics of vertical structure could be used to develop quantitative indicators of potential habitat for biodiversity (Ferris and Humphrey 1999).

Existing ways of mapping understorey vegetation have limitations. Field-based inventory is labour intensive and unsuitable for broadscale mapping. Aerial photography is unsuitable for mapping lower vegetation layers obscured by canopy vegetation (Wallace et al. 2016). High-resolution LiDAR is effective for mapping vertical structural layers of forest canopies (Filippelli et al. 2019) but is too costly for many applications. However, it may be possible to model
understorey structure in undisturbed forests from ecologically important characteristics that are already available for many forests, such as underlying geology and the landform properties.

Topography affects microclimate, drainage, soil formation, and wildfire, which in turn influence the distribution of plant species composition and vegetation structures (Reid et al. 2005). Although there are general trends in vegetation patterning in relation to topography, it is unknown whether topographic attributes relate sufficiently strongly to vegetation structure to predict local-scale habitat conditions (Fu et al. 2004). Topographic characteristics can be directly extracted from digital terrain models (DTMs) (e.g. Franklin 1998; Wilson et al. 2007). DTM s of various resolutions are widely available and can also be derived directly from LiDAR elevation datasets across a range of horizontal resolutions (Anderson et al. 2006; Jayathunga et al. 2018).

Underlying geology, especially rock type, affects vegetation most particularly through its impacts on soil characteristics, including the availability of soil nutrients and water-holding characteristics. Soil nutrients are important drivers of understorey vegetation species composition, structure, and growth (Kasel et al. 2017). To help with using geology as a predictor of forest structure, there is good quality mapping of underlying rock types for many parts of the world.

This study tests how well forest understorey structure can be predicted from topographic characteristics and underlying rock type in an economically important tall forest ecosystem. The wet eucalypt forests of Tasmania have high variability in structure over short distances and are considered to respond to topographic characteristics and underlying rock types (Reid et al. 2005). Large areas of these
forests are managed for wood and pulp production, but the understoreys of these forests harbour high levels of native biodiversity that need to be managed sustainably (Baker et al. 2016). These forests have moderate to high canopy density meaning that new tools are needed for assessing understorey structure. Specifically, this paper aims to predict the density of three canopy layers of a wet eucalypt primary forest using topographic attributes derived from a DTM plus geological data and explores the relative importance of topography and geology in influencing vegetation structure. We use high-density airborne LiDAR point clouds to derive reference datasets for the densities of the three canopy layers. We also explore the influence that the spatial resolution of the DTM has on predictive accuracy.

**Methods**

**Study site and remote sensing datasets**

This study was conducted within a topographically and geologically complex 5 km by 5 km area centred at 43.104° S 146.656° E. The area is part of the Warra Supersite within the Wilderness World Heritage Area in Tasmania, Australia, and is in the cool, temperate wet forest biome – one of the most productive terrestrial ecosystems in the world. The Supersite is a member of the Terrestrial Ecosystem Research Network (TERN) and Australian Supersite Network (TERN 2017). Mean annual rainfall is 1,707 mm and mean daily temperature ranges are 8.3°C to 19.3°C in January (summer) and 2.5°C to 8.6°C in July (winter) (Bureau of Meteorology 2017).
A 19.09 km² core area of mature native forest was used for analysis after excluding 2.79 km² of roads, rivers, and previously harvested sites and 3.12 km² of edges (Fig. 1). We utilised airborne LiDAR and a geology data layer. High-resolution LiDAR data was collected with a Riegl LMS-Q560 sensor scanning at 1064 nm from a flying height of 500 m above terrain on May 30, 2014, resulting in an average point density of 28.66 points/m² (spacing = 0.19 m).

Underlying rock types were extracted from a geology data layer based on 1:25,000 scale mapping (Mineral Resources Tasmania 2019). Rock types were generalised to avoid ambiguity in classification or because some types were both rare and geologically closely allied to other types. Thus, all Permian sedimentary rocks were integrated into a single category; the carbonaceous rock types Ordovician limestones and Cambrian dolomite were combined; Dolerite boulders were merged with Dolerite talus; Quaternary sediments were merged with Quaternary alluvium.
Maps of the study site showing the 30 m resolution digital terrain model (left) and geology (right). White areas removed from the analysis were roads, rivers, and previously harvested sites.

Abbreviations of geology types are Quaternary alluvium (Alluvium), Cambrian siliceous sediments (Cambrian), Jurassic dolerite (Dolerite), Neoproterozoic dolomite (Dolomite), Glacial tills (Glacial), Permain sediments (Permian), and dolerite talus (Talus).

This study followed the overall workflow presented in Fig. 2. Three response variables, i.e. the canopy density of a lower layer (≥2 m to ≤10 m), middle layer (>10 m to ≤30 m), and an upper layer (>30 m to ≤50 m), were derived directly from the LiDAR data. We also used the LiDAR data to produce a series of DTMs (1 m, 5 m, 10 m, 20 m, and 30 m) of varying spatial resolution from which we derived twelve topographic attributes. These two LiDAR-derived sets of data are largely independent of each other because the former is a function of returns from the bare ground and forest floor, and the latter is primarily a function of returns from middle layers. The same LiDAR data were used to derive both predictor and
response variables since the high-resolution LiDAR data will produce a ‘best case scenario’ for quantifying relationships between topography and forest structure and enables us to compare DTM resolutions without introducing other biases. Assessing the predictive capacity of other freely available DTMs was beyond the scope of this study. The topography and geology datasets provided thirteen predictor variables, which were deployed for random forest regression modelling to determine whether topography and geology predict local-scale forest structure.

![Workflow for testing the capacity for topography and geology to predict the density of vegetation in three canopy layers of a mature forest landscape.](image)

**Fig. 2** Workflow for testing the capacity for topography and geology to predict the density of vegetation in three canopy layers of a mature forest landscape.
Preparing data layers

We processed airborne LiDAR data using LAStools (Academic version 171030). Firstly, LiDAR point clouds were classified into the ground and non-ground classes using *lasground*, then normalized using *lasheight*. Outlier points (above the canopy and below ground level) were removed.

The structure of understorey layers was characterised using *lascanopy* for five different spatial resolutions (1 m, 5 m, 10 m, 20 m, and 30 m). The structure of each layer was characterised as counts of points falling into the height intervals divided by the total number of points within each grid cell. Field data from sample plots and field experience indicate that most emergent overstorey trees were above ~50 m in height, so the three understorey layers were classified as a lower layer (≥ 2 to ≤ 10 m), middle layer (> 10 to ≤ 30 m), and an upper layer (> 30 to ≤ 50 m). The canopy at the Warra silviculture system trials (far north-east of landscape) had canopy height around 40 m on poorly drained soils. Returns below 2 m height were discarded from the analysis, thereby avoiding small shrubs and coarse woody debris (Ehlers et al. 2018; Packalén and Maltamo 2006; Wilkes et al. 2016). A large proportion of forest understorey lies between 2 m and 10 m above ground (Musk 2017).

Generating topographic attributes

DTMs were created from the LiDAR data using *blast2dem* (e.g. Fig. 1). For each DTM resolution, we used SAGA-GIS (Conrad et al. 2015) to create twelve topographic attributes (Table 1, Fig. 3) considered likely to have a relationship with vegetation composition and structure (Wang et al. 2015).
Table 1 Topographic attributes and their description

| Topographic attribute                      | Description                                                                                                                                 |
|--------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Slope                                      | Calculated by fitting a plane to the eight neighbouring cells (Travis et al. 1975).                                                          |
| Aspect                                     | The orientation of the cell relative to the north (Travis et al. 1975).                                                                       |
| Catchment area                             | The upstream area of each cell (Kiss 2004).                                                                                                  |
| Profile curvature                          | The rate of change of slope in a downslope direction: a proxy for acceleration and deceleration of water over the terrain (Wilson et al. 2007). |
| Plan curvature                             | The curvature of a contour at the central pixel. It can be used as a proxy for convergence and divergence of water (Wilson et al. 2007).       |
| LS (slope length and steepness) factor     | Calculated using the upslope catchment area of each cell and the grid cell slope. This metric is used as a proxy for erosivity (Desmet and Govers 1996; Kinnell 2005). |
| Potential solar radiation ratio            | The ratio of the potential solar radiation on a sloping surface to that on a horizontal surface (Moore et al. 1991).                          |
| Topographic Position Index                 | Whether any particular pixel forms part of a positive (e.g., crest) or negative (e.g., trough) feature of the surrounding terrain (Wilson et al. 2007). |
| Terrain Ruggedness Index                   | The sum change in elevation between a grid cell and its eight neighbouring grid cells (Riley et al. 1999).                                    |
| Stream Power Index                         | A measure of the erosive power of flowing water (Jacoby et al. 2011).                                                                       |
| Compound Topographic Wetness Index         | A measure of soil moisture potential that combines contextual and site information and is used to identify potential locations of ephemeral gullies (Casali et al. 2016; Gessler et al. 1995). |
| Convergence index                         | The average bias of the slope directions of the adjacent cell from the direction of the central cell minus 90 degrees (Kiss 2004).             |
Fig. 3 Maps of the twelve topographic attributes extracted from DTMs.
Predictive modelling

We used random forest (RF) non-parametric machine learning regression modelling to predict each of the three understorey layers with the twelve topographic attributes and the geology vector data as predictor variables. RF is extensively used in ecological modelling and remote sensing studies and has the particular advantage of providing good information on the contribution of individual predictors to models (Cutler et al. 2007). The RF model training, prediction, and validation processes were implemented in R version 3.4.1 (R Core Team 2017) using libraries randomForest (Liaw and Wiener 2002), caret (Kuhn 2017), caTools (Tuszynski 2020), and pdp (Greenwell 2017). To achieve stable results and test the sensitivity of the models, 100 iterations were run (Breiman 1996; Qian et al. 2016; Yang et al. 2017) for training and testing, and the predicted models were validated. For the selection of variables and final performance statistics, average results were presented.

Model training and validation datasets

For each height layer of the understorey and each spatial resolution, values were extracted for each of the variables for 8,300 random point locations drawn (Criminisi et al. 2011; Wood et al. 2011) from the study site (excluding masked areas). In each case, these sample points were divided into a model training dataset (50%) and a validation dataset (50%). From the training dataset, 70% of sampled points were randomly drawn for training, and 30% for cross-validation (Abdel-Rahman et al. 2013; Kemppinen et al. 2018).
Tuning parameters (mtry and ntree)

To tune the RF model for optimal performance, two hyperparameters were tested—mtry (the number of variables used in each replicate run) and ntree (the total number of independent trees) (Cutler et al. 2012; Liaw and Wiener 2002; Tyralis and Papacharalampous 2017). In this study, the RF parameter optimisation was run for the 10 m resolution dataset. We considered five possible values for mtry: 3, 4, 6, 9, and 13 from the thirteen variables. The mtry value of 13 consistently yielded the lowest out-of-bag (OOB) errors and was used for subsequent modelling (Freeman et al. 2015). Predictions were made for the training dataset of the selected ntree values from 200 to 10,000. We found that the ntree parameter did not differ significantly with increasing ntree values of >500, the default in many RF regression modelling studies (Turner et al. 2018). Thus ntree of 500 was used.

Model accuracy assessment

We used root mean square error (RMSE) (Gao et al. 2018; Rocha et al. 2018) to assess model training and cross-validation, applying 100 iterations and averaged these to present model performance. Finally, the predicted models were validated with the independent dataset to test the robustness and stability of the model.

Contributions of individual variables

Variable importance (the relevance of the variables for predicting understorey layers using topographic variables) was assessed using percentage increase in mean square error (%IncMSE), one of the most widely used scores of importance
(Genuer et al. 2015). The higher %IncMSE, the more important is the variable. To show the marginal effect of predictor variables on each response variable (Friedman 2001), partial dependence plots (pdp) for the five predictor variables with the highest variable importance scores were created using the pdp package (Greenwell 2017) in R.

## Results

### Random forest model prediction and cross-validation

For all three canopy layers, models with the coarsest spatial resolution (30 m pixel size) outperformed those at smaller spatial resolutions (1 m, 5 m, 10 m, and 20 m) for both prediction and validation (Table 2). We were unable to explain the pattern of increased error for 10 m pixel size in the lower vegetation layer. The best model accuracy from the validation dataset was for the lower layer and the lowest accuracy for the middle vegetation layer. This contrasts with the prediction dataset, with the best accuracy for the upper layer, although there was little difference in the magnitude with RMSE ranging from 7.61% ($R^2 = 0.82$) for the upper layer through to 9.76% ($R^2 = 0.77$) for the middle layer. Hereafter we present results only for the 30 m resolution datasets.

| Canopy layers | RMSE (%) for five spatial resolutions |
|---------------|---------------------------------------|
|               | 1 m   | 5 m    | 10 m  | 20 m  | 30 m  |
| Upper layer   |       |        |       |       |       |
| Prediction    | 17.82 | 14.40  | 10.84 | 9.04  | 7.61  |
|               | Prediction  | Validation  |
|---------------|-------------|-------------|
| **Middle layer** |             |             |
| Prediction    | 20.81       | 16.78       |
| Validation    | 32.97       | 25.23       |
| **Lower layer** |             |             |
| Prediction    | 25.92       | 16.80       |
| Validation    | 25.61       | 18.40       |

Predictive maps for the density of vegetation in the three canopy layers show reasonable accordance with the actual densities measured with the high-resolution LiDAR dataset (Fig. 4).
Fig. 4 Maps comparing the density of forest understorey layers that were observed with high-resolution LiDAR data (left) with those predicted based on topography and geology (right) using 30 m resolution. Roads, rivers, and harvested areas appear as white.
Individual predictor variables

Geology was the most important variable for predicting all layers (Fig. 5). Sites on Jurassic dolerite geology class had dense lower and middle layers and the most sparse upper layer. The Jurassic dolerite was confined to a small area of high elevation within the study site (Fig. 1). By contrast, the Permian sediments geology class had a dense upper layer, and the least dense lower layer (Fig. 6).

Solar radiation was the most important predictor for the middle layer, and related variables (topographic position index and aspect) were also important for the upper layer (Fig. 5). Sites with high solar radiation had a dense lower layer and sparse middle layer and intermediate values for the upper layer. Sites with high topographic position index (TPI) had dense lower and middle layers and sparse upper layer. Vegetation density was greatest for the Northwest (292.5° to 337.5°) aspects for the lower canopy layer, for Southeast aspect (112.5° to 157.5°) for the middle layer, and East-Northeast (22.5° to 67.5°) aspects for the upper-layer. Wet sites (high SAGA wetness index) had dense upper layers and dry sites had dense lower layers. SAGA wetness index (SWI) increased with the increase in the SWI value from 2 to 6 which contrasted with the lower- and middle-layer models for which low SWI values had maximum influence.
Fig. 5 Variable importance scores (percentage increase in mean square error, %IncMSE if the variable is eliminated from models) for topographic and geological attributes used in predicting the density of three vegetation canopy layers at 30 m spatial resolution. (Acronyms: Terrain Ruggedness Index (TRI), Topographic Position Index (TPI), Stream Power Index (SPI), SAGA Wetness Index (SWI), and Convergence index (CI)).
Fig. 6 Partial dependence plots of solar radiation, terrain position index (TPI), aspect, and SAGA wetness index (SWI) for random forest models of the three forest layers. Higher partial dependence values indicate a greater influence of the variables on the model.
Fig. 7 Partial dependence plots of geology. The acronyms used are Quaternary alluvium (Alluvium), Cambrian siliceous sediments (Cambrian), Jurassic dolerite (Dolerite), Neoproterozoic dolomite (Dolomite), Glacial tills (Glacial), Permian sediments (Permian), and Dolerite talus (Talus). All Permian sedimentary rocks were integrated into a single category; the carbonaceous rock types Ordovician limestones and Cambrian dolomite were combined; Dolerite boulders were merged with Dolerite talus; Quaternary sediments were merged with Quaternary alluvium.

Discussion

We found that the density of vegetation of three understorey canopy layers of a mature forest landscape could be predicted with reasonable accuracy from a geology layer and twelve topographic indices derived from a DTM. Of five spatial resolutions tested, the best model performance for all three canopy layers was achieved at 30 m pixel size.

The capacity of topography and geology to predict forest understorey structure

The relatively good predictions for the preferred 30 m spatial resolution (RMSE = 8.97% for the lower layer; RMSE = 11.55% for the upper layer; RMSE = 13.69% for the middle layer) indicate that topography and geology have direct or indirect effects on forest structure. However, other factors are likely to be important. Spatial variability in forest age, canopy closure (and resulting below-canopy light environment), and seed availability are likely to be important drivers of the size and distribution of mid-storey species (Morsdorf et al. 2010) but are unlikely to be captured in the topographic and geological predictors used here. All the forest in our study area is likely to be at least 80 years old, with the most recent wildfire
occurring in 1934 (Hickey et al. 1999). Thus, the forest was relatively mature and relatively homogenous in age, and future research could assess relationships between forest structure and topography and geology for younger forest age classes.

Our results compare well to studies measuring vegetation density directly with LiDAR. Using discrete-return LiDAR in both conifer and deciduous forest stands in Germany, Latifi et al. (2016) showed the top and the bottom layers could be predicted better than the middle layers; a comparable result to our study. Morsdorf et al. (2010) achieved an overall accuracy of 80% to 90% for dominant layers and around 48% for sub-dominant layers, and they mentioned that accuracies could be lower in more complex plots. Wing et al. (2012) found understorey vegetation cover with the accuracies ranging from 20% to 45% and projected the accuracies of 77% by combining leaf-off and leaf-on datasets in their deciduous forest ecosystem. Also, Suchar and Crookston (2010) reported adjusted $R^2$ values of 0.22 and 0.24 for the percent shrub cover models for the forest ecosystems of the north-western United States, and they argued that shrub-herb cover is more heterogeneous than overstorey cover attributes.

**Optimal spatial resolution**

Our use of a single high-resolution LiDAR dataset to derive a range of spatial resolutions for DTMs allowed us to determine that the optimal spatial resolution for predicting understorey density of different vegetation layers in our study system was approximately 30 m for all three canopy layers. Although we did not assess resolutions greater than 30 m, field knowledge indicates high spatial turnover in our forests at such scales, meaning that coarser resolutions are likely
to lead to smoothing and loss of important information. Finding the optimal resolution was particularly valuable because, as noted in other studies (e.g. Wood et al. 2011), the drop in accuracy using finer resolutions was substantial, especially for some layers. In particular, Zald et al. (2014) showed that LiDAR metrics characterised trees with reduced accuracy as the plot size decreased. Although previous studies suggest that a 10 m resolution of DTM would be sufficient for geomorphic and hydrologic modelling (Jenkins and Coops 2011; Murphy et al. 2011; Wang et al. 2011; Zhang and Montgomery 1994), our study found poor performance at this resolution for the lower understory. The better performance at a 30 m scale compared to 10 m or less is related to the crown size of emergent trees, which are generally at least 10 m in diameter in these forests. Azaele et al. (2012) found that objects (e.g. tree crowns) that occupied a pixel were measured at higher accuracy.

Influences of topographic attributes and geology on forest structure

Geology was the best predictor of vegetation structure in the lower and upper layers and the second-best predictor for the middle layer. Such links have been shown in other systems. For example, Simonson et al. (2014) predicted shrub understoreys could be more developed because of increasing soil nutrients and water availability, both of which are likely to be linked to the underlying geology. Combining some geological categories was essential for analytical reasons but inevitably led to the loss of geological information. Modelling using a larger spatial area could permit the incorporation of more detailed geological information, which could improve the predictive power of the models.
Topographic indices collectively contributed greatly to predicting vegetation structure. This may relate to the topographical complexity of the Warra Supersite forest. Two potential contributing factors are the impacts of fire and impacts on water balance. Wildfire disturbance is a major driver of vegetation patterning (Hickey et al. 1999). The topography is well known to influence fire behavior, and consequently can be a major driver of plant species patterning (Wood et al. 2011). Our study landscape included understorey species associated with two seral stages, although in practice the species co-occur in different proportions. Early successional wet sclerophyll species have adaptations for frequent fire, while rainforest species occur where the fire is less severe and/or frequent (van Galen et al. 2018; Wood et al. 2011). Thus, topographic attributes impacting moisture status and fire history would likely have been important drivers of the species and structural composition of our landscape.

Solar radiation was considered the second most influential attribute in determining the vegetation structure, being in the top three for variable importance for each canopy layer, and greatest importance for the middle layer. Jenkins and Coops (2011) pointed out the importance of higher insolation levels for canopy characteristics derived from LiDAR in eucalypt forests in New South Wales, Australia. Areas with high solar radiation are likely to be warmer and drier than areas with low solar radiation. This is likely to increase the frequency and intensity of fires, and potentially may affect vegetation through differences in water demand in summer. Weisberg et al. (2003) noted that the more open the upper vegetation layers, the more light that penetrates to lower layers, which determines the positive associations with understorey species richness. Simonson et al. (2014) found less developed plant canopies due to reduced light
transmission. They found canopy cover and density were higher on north-facing aspects receiving less solar radiation.

Moisture availability is a very important determinant of plant species composition and forest structure (Bartels and Chen 2013; Campbell et al. 2018). In general, the partial dependence plots demonstrated that the lower the value of the SAGA wetness index (SWI), the higher the understorey density for the lower and middle layers. The wetness index helped to characterise the biological distribution and species diversity elsewhere (Moore et al. 1988).

**Conclusions**

This paper demonstrates that DTM-derived topographic attributes and geology information can be used to predict the understorey density of three canopy layers for a mature wet eucalypt forest landscape with moderately high accuracy. Underlying geology and a suite of topographic attributes have significant, but differing, relationships with the lower, middle, and upper understorey canopy layers.

The 30 m resolution DTM dataset was best for predicting the structure of all three levels of the understorey in our study system. However, the optimum resolution would need to be determined separately for other landscapes, ecological processes, and habitat types.

This study derived the DTM for calculating topographic indices from high-resolution LiDAR. This would be cost-prohibitive for broad-scale applications.

One key future step would therefore be to compare the predictive capacity of
topographic attributes derived from DTMs from other sources such as low-resolution LiDAR, satellite RADAR, or statewide topographic mapping from aerial photography (TASMAP eSHOP 2015). However, the strong performance of the lowest resolution DTM (30 m) makes it likely that DTMs derived from moderate and even low-resolution LiDAR may perform well.

In general, geology performed the best for predicting vegetation density of the lower and upper layers, while solar radiation, followed by geology was best for predicting the middle layer in a wet eucalypt primary forest. Topographic position index, aspect, and SAGA wetness index had moderate importance.

Overall, these results could be used to predict understorey vegetation structural layers and biomass, which can have important implications for determining fuel loads, carbon stores, and habitat quality for biodiversity. The capacity to predict forest structure from digital terrain models and geology information as demonstrated here could be broadly useful to foresters and ecologists and contribute to sustainable forest management and biodiversity conservation.

Acknowledgements

We acknowledge the Terrestrial Ecosystem Research Network (TERN) and Airborne Research Australia (ARA) for funding and collection of airborne LiDAR data. This study was partially funded by the Australian Research Council Linkage Project LP140100075 with partner organisations Sustainable Timber Tasmania and VicForests. Yadav was supported by a Tasmania Graduate Research Scholarship from the University of Tasmania.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

ORCID

Bechu K.V. Yadav: https://orcid.org/0000-0002-1076-6099

Arko Lucieer: https://orcid.org/0000-0002-9468-4516

Susan C. Baker: https://orcid.org/0000-0002-7593-0267

Gregory J. Jordan: https://orcid.org/0000-0002-6033-2766

References

Abdel-Rahman EM, Ahmed FB, Ismail R (2013) Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. Int J Remote Sens 34:712-728 doi:https://doi.org/10.1080/01431161.2012.713142

Anderson ES, Thompson JA, Crouse DA, Austin RE (2006) Horizontal resolution and data density effects on remotely sensed LIDAR-based DEM. Geoderma 132:406-415 doi:https://doi.org/10.1016/j.geoderma.2005.06.004

Azaele S, Cornell SJ, Kunin WE (2012) Downscaling species occupancy from coarse spatial scales. Ecol Appl 22:1004-1014 doi:https://doi.org/10.1890/11-0536.1
Baker SC et al. (2016) A cross-continental comparison of plant and beetle responses to retention of forest patches during timber harvest. Ecol Appl 26:2495-2506. doi: https://doi.org/10.1002/eap.1406

Bartels SF, Chen HYH (2013) Interactions between overstorey and understory vegetation along an overstorey compositional gradient. J Veg Sci 24:543-552. doi: https://doi.org/10.1111/j.1654-1103.2012.01479.x

Breiman L (1996) Bagging Predictors. MLear 24:123–140. doi: https://doi.org/10.1023/A:1018054314350

Bureau of Meteorology (2017) Climate statistics for Australian locations (2004-2017). Australian Government, Australia

Campbell MJ, Dennison PE, Hudak AT, Parham LM, Butler BW (2018) Quantifying understory vegetation density using small-footprint airborne LiDAR. Remote Sens Environ 215:330-342. doi: https://doi.org/10.1016/j.rse.2018.06.023

Camprodon J, Brotons L (2006) Effects of undergrowth clearing on the bird communities of the Northwestern Mediterranean Copice Holm oak forests. For Ecol Manage 221:72–82. doi: https://doi.org/10.1016/j.foreco.2005.10.044

Casalí J, Chahor Y, Giménez R, Campo-Bescós M (2011) A critical discussion on the applicability of Compound Topographic Index (CTI) for predicting ephemeral gully erosion. In: Geophysical Research Abstracts, Vienna, Austria, 2016. vol EGU2016-15091. EGU General Assembly, held 17–22 April, 2016.

Conrad O et al. (2015) System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci Model Dev 8:1991-2007. doi: https://doi.org/10.5194/gmd-8-1991-2015

Crinimisi A, Shotton J, Konukoglu E (2011) Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Microsoft Research Ltd, 7 JJ Thomson Ave, Cambridge, CB3 0FB, UK, Cambridge, CB3 0FB, UK

Cutler A, Cutler DR, Stevens JR (2012) Random Forests. In: Zhang C, Ma Y (eds) Ensemble Machine Learning: Methods and Applications. Springer US, Boston, MA, pp 157-175. doi: https://doi.org/10.1007/978-1-4419-9326-7_5

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson JC (2007) Random forests for classification in ecology. Ecology 88:2783-2792. doi: https://doi.org/10.1890/07-0539.1

Desmet PJJ, Govers G (1996) A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units. J Soil Water Conserv 51:427-433

Ehlers S et al. (2018) Assessing Error Correlations in Remote Sensing-Based Estimates of Forest Attributes for Improved Composite Estimation Remote Sens 10:667

Ferris R, Humphrey JW (1999) A review of potential biodiversity indicators for application in British forests. Forestry: An International Journal of Forest Research 72:313–328. doi: https://doi.org/10.1093/forestry/72.4.313

Filippelli SK, Lefsky MA, Rocca ME (2019) Comparison and integration of LiDAR and photogrammetric point clouds for mapping pre-fire forest structure. Remote Sens Environ 224:154-166. doi: https://doi.org/10.1016/j.rse.2019.01.029

Franklin J (1998) Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. J Veg Sci 9:733-748. doi: https://doi.org/10.2307/3237291

Freeman EA, Moisen GG, Coulston JW, Wilson BT (2015) Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. Can J For Res 46:323-339. doi: https://doi.org/10.1139/cjfr-2014-0562

Friedman JH (2001) Greedy Function Approximation: A Gradient Boosting Machine. Ann Stat 29:1189–1232

Fu BJ, Liu SL, Ma KM, Zhu YG (2004) Relationships between soil characteristics, topography and plant diversity in a heterogeneous deciduous broad-leaved forest near Beijing, China. Plant Soil 261:47-54. doi: https://doi.org/10.1023/B:PLSO.0000035567.97093.48

Gao Y, Lu D, Li G, Wang G, Chen Q, Liu L, Li D (2018) Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region. Remote Sens 10:627. doi: https://doi.org/10.3390/rs10040627

Genuer R, Poggi J-M, Tuleau-Malot C (2015) VSURF: An R package for variable selection using random forests. The R Journal 7:19-33.
Gessler PE, Moore ID, McKenzie NJ, Ryan PJ (1995) Soil-landscape modelling and spatial prediction of soil attributes. Int J Geogr Inf Syst 9:421-432 doi:https://doi.org/10.1080/02693799508902047

Greenwell BM (2017) pdp: An R package for constructing partial dependence plots. The R Journal (contributed research article) 9/1:421-436

Hickey JE, Su W, Rowe P, Brown MJ, Edwards L (1999) Fire history of the tall wet eucalypt forests of the Warra ecological research site, Tasmania. Aust For 62:66-71 doi:https://doi.org/10.1002/1059-0597(199910)62:3<66::AID-AF1>3.0.CO;2-H

Jacoby BS, Peterson EW, Dogwiler T (2011) Identifying the Stream Erosion Potential of Cave Levels in Carter Cave State Resort Park, Kentucky, USA J Geogr Inf Syst 3:323-333 doi:10.4236/jgis.2011.34030

Jayathunga S, Owari T, Tsuyuki S (2018) Evaluating the performance of photogrammetric products using fixed-wing UAV imagery over a mixed conifer–broadleaf forest: Comparison with airborne laser scanning. Remote Sens 10:1 doi:10.3390/rs10020187

Jenkins RB, Coops NC (2011) Landscape controls on structural variation in eucalypt vegetation communities: Woronora Plateau, Australia. Aust Geogr 42:1-17 doi:https://doi.org/10.1080/00049182.2011.546316

Kasel S, Bennett LT, Aponte C, Fedrigo M, Nitschke CR (2017) Environmental heterogeneity promotes floristic turnover in temperate forests of south-eastern Australia more than dispersal limitation and disturbance. Landsc Ecol 32:1613-1629 doi:https://doi.org/10.1007/s10980-017-0526-7

Kemppinen J, Niittynen P, Riihimäki H, Luoto M (2018) Modelling soil moisture in a high-latitude landscape using LiDAR and soil data. Earth Surf Process Landf 43:1019-1031 doi:https://doi.org/10.1002/esp.4301

Kinnell P (2005) Alternative approaches for determining the USLE-M slope length factor for grid cells. Soil Sci Soc Am J 69:674-680 doi:https://doi.org/10.2136/sssaj2004.0047

Kiss R (2004) Determination of drainage network in digital elevation models, utilities and limitations. J Hung Geomat 2:16-29

Kuhn M (2017) Classification and Regression Training (Package ‘caret’).

Latifi H, Hill S, Schumann B, Heurich M, Dech S (2017) Multi-model estimation of understorey shrub, herb and moss cover in temperate forest stands by laser scanner data. Forestry: An International Journal of Forest Research 90:496-514 doi:https://doi.org/10.1093/forestry/cpw066

Liaw A, Wiener M (2002) Classification and Regression by randomForest. vol 2/3.

Mineral Resources Tasmania (2019) Digital Geological Atlas 1:25,000 Scale Series. Tasmanian Government, Department of State Growth.

Murphy PNC, Ogilvie J, Meng F-R, White B, Bhatti JS, Arp PA (2011) Modelling and mapping topographic variations in forest soils at high resolution: A case study. Ecol Modell 222:2314-2332 doi:https://doi.org/10.1016/j.ecolmodel.2011.01.003
Wilkes P et al. (2016) Using discrete-return airborne laser scanning to quantify number of canopy strata across diverse forest types. Methods Ecol Evol 7:700-712. doi: https://doi.org/10.1111/2041-210X.12510

Wilson MFJ, O’Connell B, Brown C, Guinan JC, Grehan AJ (2007) Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. Mar Geodesy 30:3-35 doi: https://doi.org/10.1080/01490410701295962

Wing BM, Ritchie MW, Boston K, Cohen WB, Gitelman A, Olsen MJ (2012) Prediction of understory vegetation cover with airborne lidar in an interior ponderosa pine forest. Remote Sens Environ 124:730-741 doi: https://doi.org/10.1016/j.rse.2012.06.024

Wood SW, Murphy BP, Bowman DMJS (2011) Firescape ecology: how topography determines the contrasting distribution of fire and rain forest in the south-west of the Tasmanian Wilderness World Heritage Area. J Biogeogr 38:1807-1820 doi: https://doi.org/10.1111/j.1365-2699.2011.02524.x

Yang Y, Qian W, Zou H (2017) Insurance premium prediction via gradient tree-boosted tweedie compound poisson models. Bus Econ Stat:1-15 doi: https://doi.org/10.1080/07350015.2016.1200981

Zald HSJ, Ohmann JL, Roberts HM, Gregory MJ, Henderson EB, McGaughey RJ, Braaten J (2014) Influence of LiDAR, Landsat imagery, disturbance history, plot location accuracy, and plot size on accuracy of imputation maps of forest composition and structure. Remote Sens Environ 143:26-38 doi: https://doi.org/10.1016/j.rse.2013.12.013

Zhang W, Montgomery DR (1994) Digital elevation model grid size, landscape representation, and hydrologic simulations. Water Resour Res 30:1019-1028 doi: https://doi.org/10.1029/93WR03553