

Article

The Performance of Random Forests in an Operational Setting for Large Area Sclerophyll Forest Classification

Andrew Mellor ^{1,2,3*}, Andrew Haywood ^{2,3}, Christine Stone ⁴ and Simon Jones ¹

¹ School of Mathematical and Geospatial Sciences, RMIT University, GPO Box 2476, Melbourne, VIC 3001, Australia; E-Mail: simon.jones@rmit.edu.au

² Victorian Department of Environment and Primary Industries, 8 Nicholson Street, East Melbourne, VIC 3002, Australia; E-Mail: andrew.haywood@dse.vic.gov.au

³ Joint Remote Sensing Research Program, School of Geography, Planning and Environmental Management, University of Queensland, St Lucia, QLD 4072, Australia

⁴ New South Wales Department of Primary Industries, P.O. Box 100, Beecroft, NSW 2119, Australia; E-Mail: christine.stone@dpi.nsw.gov.au

* Author to whom correspondence should be addressed; E-Mail: andrew.mellor@rmit.edu.au; Tel.: +61-3-9925-2283; Fax: +61-3-9925-2454.

Received: 17 April 2013; in revised form: 10 May 2013 / Accepted: 25 May 2013 /

Published: 4 June 2013

Abstract: Mapping and monitoring forest extent is a common requirement of regional forest inventories and public land natural resource management, including in Australia. The state of Victoria, Australia, has approximately 7.2 million hectares of mostly forested public land, comprising ecosystems that present a diverse range of forest structures, composition and condition. In this paper, we evaluate the performance of the Random Forest (RF) classifier, an ensemble learning algorithm that has recently shown promise using multi-spectral satellite sensor imagery for large area feature classification. The RF algorithm was applied using selected Landsat Thematic Mapper (TM) imagery metrics and auxiliary terrain and climatic variables, while the reference data was manually extracted from systematically distributed plots of sample aerial photography and used for training (75%) and accuracy (25%) assessment. The RF algorithm yielded an overall accuracy of 96% and a Kappa statistic of 0.91 (confidence interval (CI) 0.909–0.919) for the forest/non-forest classification model, given a Kappa maximised binary threshold value of 0.5. The area under the receiver operating characteristic plot produced a score of 0.91, also indicating high model performance. The framework described in this study contributes to the operational deployment of a robust, but affordable, program, able to collate and process

large volumes of multi-sourced data using open-source software for the production of consistent and accurate forest cover maps across the full spectrum of Victorian sclerophyll forest types.

Keywords: large area monitoring; forest extent; random forests; operational; Landsat TM; MODIS

1. Introduction

Forest extent is a measure commonly assessed in national forest inventories (NFI) [1] and, under the Montreal process [2], is a specific indicator used for monitoring and reporting sustainable forest management. For natural resource management agencies, current and accurate forest area estimates are critical for effective environmental monitoring. While ground-based (field plot) forest inventories provide accurate and unbiased forest area estimates, spatially explicit remote sensing-derived forest extent maps can be used to assess the spatial configuration of forest at the landscape scale and used in combination with a high resolution sample (two-staged sampling) to improve forest area estimates [3].

In Australia, under the Australian National Forest Inventory, forest is defined as “A land area, incorporating all living and non-living components, dominated by trees having usually a single stem and a mature or potentially mature stand height exceeding two metres and with existing or potential crown cover of overstory strata about equal to or greater than 20 percent. This definition includes native forests and plantations and areas of trees that are sometimes described as woodlands” [4]. The structural components in this definition encompass a wide range of forest types, from open low sparse canopy woodland to tall dense canopy forests (as illustrated by Figure 1, [5]).

In Australia (and the state of Victoria, in particular), dry, damp and wet sclerophyll forests and woodlands comprise many of the forested ecosystems. The canopies in these ecosystems are dominated by eucalypt species and are characteristically open with irregular (asymmetrical) crown configurations and low foliage density [6]. Canopy foliage is often clumped, leaves tend to concentrate around crown perimeters [7] and exhibit an erectophile (vertical) leaf angle distribution. In Victoria, as in much of Australia’s forests, there is a high diversity of forest development phases, vertical and horizontal forest structures, topography and soil types [8], as well as dynamic phenological processes in understory vegetation [9].

These characteristics pose a number of challenges to the use of remote sensing in these environments for classifying and mapping forests. The mid- and under-story components, shadows and background soils all exhibit a strong influence on spectral reflectance characteristics. From a synoptic perspective, forest cover in Victoria can appear indistinguishable from shrub and other low and sparse woody vegetation species. Complexity and background noise in remote sensing signatures from open sclerophyll eucalypt forests is further intensified by the influence of dynamic understory elements and variation in forest structures [10]. The challenges and complexities associated with forest extent mapping across state and territories in Australia is evidenced by large differences and inconsistencies in forest extent maps and forest area estimates produced by state and federal government agencies and the variability in forest area estimates published in Australia’s national five-yearly State of the Forests

reports [11]. The processing of large area remote sensing datasets poses a further challenge for state land management agencies.

Figure 1. Australian forest structural definitions [5].



Source: Australian Land Information Group and JA Carnahan (1990). *Atlas of Australian Resources, Vegetation*. Australian Government Publishing Service, Canberra.

Random Forests (RF) [12] offers a possible solution to address these large area forest classification challenges, universal across many of Australia's forest ecosystems. Machine learning classifiers, such as RF, are increasingly being used for environmental mapping and modelling applications in fields, such as natural resource management and forestry [13–15]. RF is an ensemble decision tree classifier, which combines bootstrap sampling to construct many individual decision trees, from which a final class assignment is determined [12].

RF can be used to learn complex non-linear relationships, such as those present in variable vertical forest structure and the association of overstorey to understorey forest vegetation. RF has been demonstrated to be very effective for accurate land cover mapping across complex and heterogeneous landscapes [15] and to be relatively insensitive to noise [15], making it suitable for application in complex and dynamic forest environments. As RF does not require normally distributed model training data, its application is appropriate for areas where species distributions of ecological communities follow non-linear patterns across the landscape [16] and where complex terrain effects data normality [17]. Other reported benefits of RF include its relative insensitivity to outliers [12,18], common characteristics of open canopies across large areas of dynamic and highly variable forest ecosystems. Furthermore, the RF classifier runs efficiently on large datasets [15], making it suitable for regional-scale mapping, comprising millions of hectares.

As only a random subset of variable data is used to construct each decision tree in a random forest classifier ensemble, correlation between decision trees is reduced, thereby improving predictive power and classification accuracy, whilst decreasing the computational complexity of the algorithm. As has been demonstrated in recent studies [19–22], RF can incorporate multiple-sources of remote sensing data with ancillary continuous and categorical biophysical spatial data to improve classification performance and discriminate between forest and non-forest.

Moderate resolution multi-spectral imagery, such as Landsat Thematic Mapper (TM)/Enhanced Thematic Mapper (ETM+) has been commonly applied for estimating forest cover [23,24], discrimination of some forest types [25], forest cover change detection [26,27] and for model-based forest area estimation [1]. Because of the challenges described above, limitations arise in classifying forest extent where different forest structures and composition and land cover types can appear spectrally alike using traditional remote sensing data analysis techniques. Improved forest classification accuracy and forest area estimates have been achieved for large areas using multi-temporal imagery, e.g., MODIS [28,29]. The high temporal resolution of the MODIS sensor can provide valuable information about the phenological variability of different land covers and, as such, help address the challenge of forest canopy-to-understorey discrimination in the type of open canopy forest environments described above.

In the context of open-canopy forest extent classification, textural information (spatial variation data derived from optical imagery) can provide additional information to a RF classifier, by differentiating vegetation that appears spectrally similar when integrated into a remote sensing image pixel, but whose spatial patterns differ [30]. Recent studies have used satellite image-derived texture indices to improve forest stand classification [31], biomass and carbon estimation [25,32] and forest structure derivation [33]. In a large heterogeneous landscape RF classification study, Rodríguez-Galiano *et al.* [34] increased overall accuracy by 8% (and Kappa by 9%) through the inclusion of textural information.

The conditional relationships between forest vegetation and biophysical factors can also be used to further improve forest/non-forest discrimination. Species-environment relationships are central to predictive geographical modelling [35]. Topographic variables (e.g., elevation, slope and aspect) used in combination with spectral data have been demonstrated to enhance forest, habitat and vegetation classification [19–22]. Bioclimatic maps (e.g., temperature, precipitation) are an additional source of commonly used ancillary classification data. These maps are typically developed using elevation-sensitive interpolation of climate station data and digital elevation models [35], which support the assumption that climate has a major influence on species distribution at broad geographic scales [36] and that similar compositions of vegetation can be expected to occur at sites with comparable soil, climate and topography [37]. In this paper, we evaluate the operational performance and utility of RF for classifying forest extent across Victoria, Australia, using remote sensing, topographic and climate predictor variables. The originality of this study lies firstly in the scale of the application of the RF algorithm, to construct, evaluate and implement an RF classifier to produce an accurate ~220,000 km² land management agency forest map. As far as we know, this scale of RF operation is unique. The second novel aspect to this study is in its application setting, which, to our knowledge, is the first time RF has been used in an operational environment at a regional scale comprising highly diverse and complex Australian forest ecosystems and topography, dominated by open canopy sclerophyll forests and woodland.

While studies on the production of forest and land cover maps derived from RF (or similar) classification techniques using multi-source remote sensing and ancillary data are published routinely in the academic literature, a secondary objective of this paper is to describe a framework for operational implementation of the RF algorithm using open-source software. The framework includes each phase of the RF classification process (from predictor variable pre-processing, through model development and implementation), to support transfer of this technology in an operational land management agency context and make use of the freely available and growing archives of remote sensing and geographic data.

2. Random Forests

Random Forests uses bootstrap aggregated sampling (bagging) to construct many individual decision trees from which a final class assignment is determined [18]. The RF algorithm constructs each decision tree using a bootstrap sample from available training data, with the remaining assigned as out-of-bag (OOB) samples. At each decision tree node, a random subset of predictor variables are tested to partition the observation data into increasingly homogeneous subsets. The node-splitting variable selected from the variable subset is that which results in the greatest increase in data purity (variance or Gini) before and after the tree node split [18]. Tree building continues until there are no further gains in purity. A response variable can be predicted as an average (continuous variable classification) or model vote (categorical classification) among all decision trees built in the forest. The OOB sample data are used to compute accuracies and error rates averaged over all predictions [18] and estimate variable importance in the classification. The computational complexity of the algorithm is reduced, as only a random subset of variables is used at each node split. This process also reduces correlation between trees, thereby improving both predictive power and classification accuracy. RF includes two methods to estimate the importance of each predictor variable in the model. The mean decrease in accuracy (MDA) importance measure is calculated as the normalised difference between OOB accuracy of the original observations to randomly permuted variables [18]. An alternative variable importance measure is calculated by summing all of the decreases in Gini impurity at each tree node split, normalised by the number of trees [38,39].

3. Open-Source Software

By adopting an open-source framework for spatial data management, processing and analysis, users, such as land management agencies, can benefit from freely available software products and access to source code through which new algorithms can be integrated and manipulated. Stallman [40] describes the four freedoms of the free and open-source software approach, as freedom to (i) run the program for any purpose, (ii) study how the program works, (iii) redistribute copies and (iv) improve the program and release such improvements to the public [41].

3.1. Geographic Resources Analysis Support System (GRASS)

GRASS (Geographic Resources Analysis Support System) [42] is an open-source geographical information system capable of handling raster, topological vector, image processing and graphic data.

Released under the GNU General Public License (GPL), GRASS is developed by a multi-national group of developers and is one of the eight initial software projects of the Open Source Geospatial Foundation. GRASS has a modular structure into which may be plugged new routines programmed in a variety of languages (e.g., Python, C, shell), and there are over 300 modules and more than 100 add-on modules for the creation, manipulation and visualisation of both raster and vector data. The GRASS modules are designed under the UNIX philosophy (*i.e.*, that programs work together and handle text streams) and can be combined using shell scripting to create more complex or specialized modules by a user. GRASS supports an extensive range of raster and vector formats through GDAL/OGR libraries, including OGC-conformal (Open Geospatial Consortium) Simple Features for interoperability with other GIS.

3.2. R and Python

R [43] is an open-source language and software environment commonly used in research fields for statistical computing and graphics. One of the main advantages of R is its object-orientated approach, which allows results of statistical procedures to be stored as objects and used as input in further computations. R is a simple and effective formal complete programming language, and the R environment is, therefore, highly extensible. GRASS and R software can be integrated through the R package, *spgrass* [44], an interface allowing GRASS GIS functions to be implemented within R code and data to be easily exchanged between the two software packages. Python [45] is an object-orientated high-level programming language that is widely used as a scripting language in the spatial analysis environment. Python's popularity has led to the creation of many useful libraries, increasing its flexibility and interoperability, and it has well developed modules for linking with GRASS and R.

4. Methods

4.1. Study Area

The study area comprises approximately 7.2 million hectares of public land forests and parks tenure (hereafter, referred to as public land forests) in the state of Victoria, in southeast Australia. This area includes 4 million ha of national parks and conservation reserves, managed primarily for ecosystem and biodiversity protection, tourism and recreation. The remaining 3.2 million ha are multiple-use state forest tenure, which include the provision of timber and non-timber forest products. Bounding extents of Victoria are north 141°47'36" E 33°58'54"S, east 149°58'36"E 37°30'20"S, south 146°17'13"E 39°9'33"S and west 140°57'29"E 34°28'23"S.

Public land forests extend to all parts of the state and range from low multi-stemmed Mallee woodland across flat and gently undulating topography in the Northwest and Box-Ironbark forests, characterised by sparse to dense canopies of box, ironbark and gum-barked eucalypts up to 25 m tall, on flat to undulating landscapes on rocky, auriferous soils across central Victoria. Highly variable medium and tall canopy damp sclerophyll forests are widespread across the study area, found on a range of loamy, clay-loam and sandy-loam soils. Tall (up to and above 75 m) wet sclerophyll forests are found mostly in the eastern part of the study area on deep loamy soils at higher elevations. Dry

sclerophyll forests are prevalent throughout the east, central and southwest parts of the study area on clay-loam, sandy-loam and shallow rocky soils of exposed hillsides, with canopies typically less than 25m tall, with crooked, spreading trees [46].

The study area is characterised by a range of different climate zones and diverse topography. The northwest region experiences semi-arid conditions, with low median annual rainfall (less than 250 mm in parts), with coastal areas experiencing a cooler temperate climate. Dry inland plains dominate much of the central and western parts of the state. The Victorian Alps—part of the Australian Great Dividing Range mountain system—extend east-west from the centre of the study area, with elevation up to 2,000 m. The Victorian Alps experience the lowest average temperatures and highest precipitation (greater than 1,400 mm/yr) in the study area. This variety of climate and topography is reflected in the variation in forest types and structure across the study area.

4.2. Training Data

Classification training data were derived from seven hundred and sixty-six 2×2 km land cover maps, systematically distributed across the Victorian Forest Monitoring Program (VFMP) [47] random stratified grid (Figure 2). On-screen digital aerial photographic interpretation (API) of high-resolution (30 cm and 50 cm pixels) colour aerial photographs (photoplots) across the study area (acquired over the period 2006 to 2010) were used to create the land cover maps, based on a land cover classification system [48] comprising broad forest type, canopy height and cover. The delineation of landscape objects into broad forest type/land cover classes, three canopy cover and three height classes, was undertaken by trained interpreters. Crown shape, size and arrangement, shadow and photographic image colour were all used for interpretation of the aerial photography. For the classification of forest, the Australian National Forest Inventory (NFI) forest definition [49] was used, with an applied 0.5 ha minimum mapping unit, consistent with the UNFAO forest definition [50].

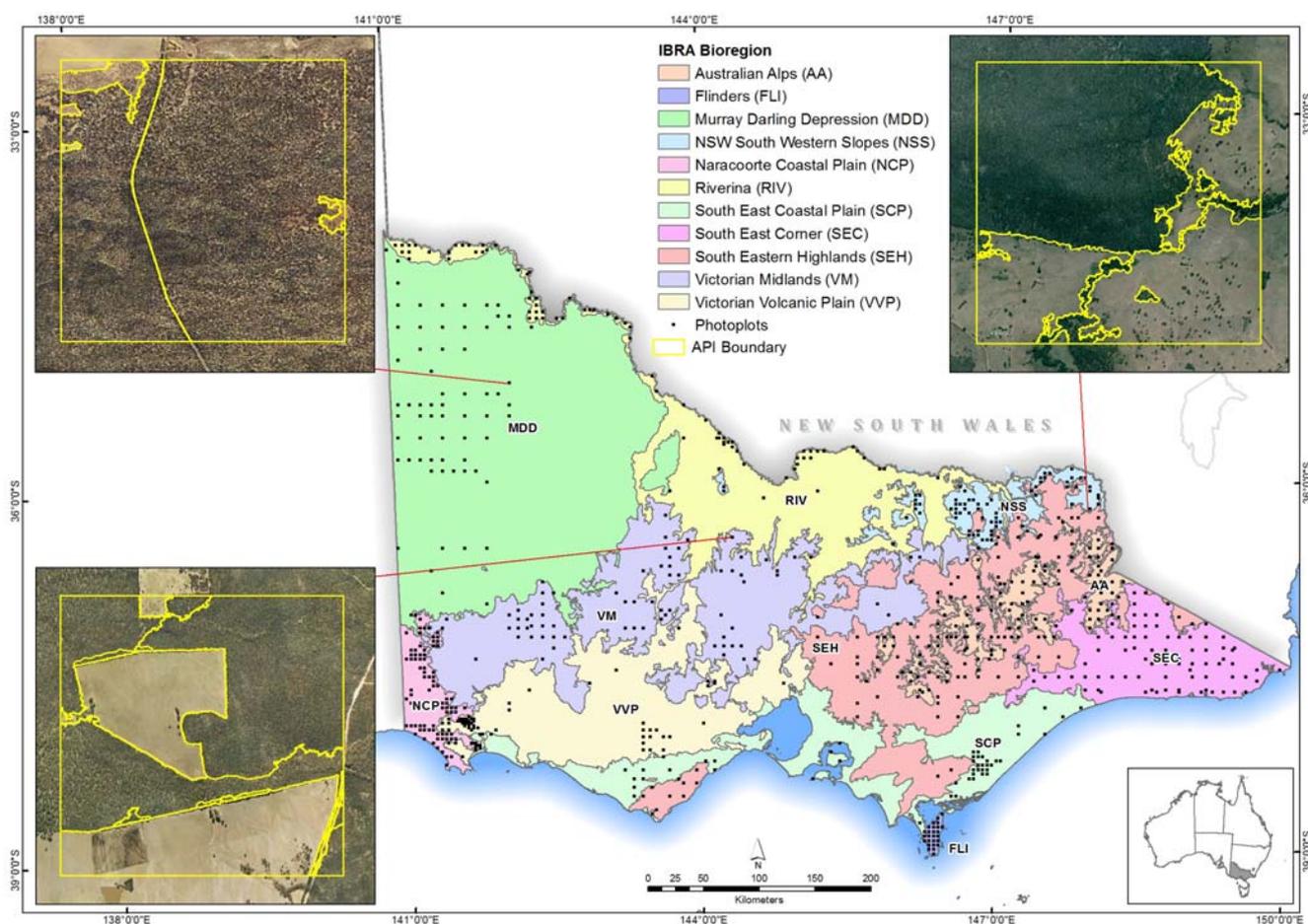
API data were aggregated into forest and non-forest training data classes. Mapping on pre- and post-2008 photography was adjusted to a baseline date of December 31, 2008, using ancillary GIS data to re-attribute and update API polygons, based on major known land cover changes associated with wildfire and clear fell logging. Training data API maps are further stratified by IBRA (Interim Biogeographic Regionalisation for Australia) Bioregions—relatively large, geographically distinct areas of land that share common characteristics, including geology, landform patterns, climate, ecological features and plant and animal communities. Eleven Bioregions are located within the study area. Figure 2 shows the distribution of VFMP sample land cover maps across the study area and Bioregions and example API land cover maps. For further information on the API method, refer to [51]. API vector data were converted to raster format to align with the 30×30 m pixels of Landsat satellite imagery (described in Section 4.3).

4.3. Predictor Variables

Nineteen cloud-free Landsat TM scenes were used to build a study area mosaic; selected and downloaded from USGS Earth Explorer [52]. Satellite images were acquired between February and March 2009, corresponding to late summer conditions with relatively high scene sun angles (to minimise shadow and terrain effects) and designed to maximise spectral differences between overstory

evergreen woody vegetation and seasonal understory vegetation. Where cloud-free images were unavailable, the acquisition period was extended to December 2008 or the summer period in the preceding or following year. Images were downloaded in USGS L1T georectified and terrain-corrected format, at a spatial accuracy considered acceptable for the study (\pm one 30 m/pixel). Landsat TM spectral bands 1–5 and 7 were pre-processed to minimise sources of between-scene spatial and temporal variation associated with different atmospheric conditions, topography, sensor location and sun elevation. A physical model was applied to convert image digital numbers (DNs) to surface reflectance standardised to a fixed viewing and illumination geometry, incorporating the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model [53], using a methodology described in [54]. Pre-processed image tiles were mosaicked to create six study area surface reflectance Landsat TM bands.

Figure 2. Victorian Interim Biogeographic Regionalisation for Australia (IBRA Bioregions) and aerial photographic interpretation (API) land cover maps.



Textural indices were derived from an NDVI layer produced using the Landsat TM surface reflectance bands 3 and 4, rescaled to a 6-bit raster (64 grey levels). Three first order (occurrence) texture measures were calculated using 3×3 , 5×5 and 7×7 cell neighbourhood moving windows across the grey-scaled [55] NDVI layer—these were variance, diversity (number of different values within the neighbourhood) and interspersion (proportion of cells in the neighbourhood, which differ from values assigned to the centre cell in the neighbourhood plus one). Three different sizes of neighbourhood windows were designed to capture the range in ecosystem textural variance across the study area.

Phenological temporal-variance in the study area was derived from state-wide multi-temporal MODIS NDVI data (MOD13Q1). A multi-temporal raster stack of twenty-three 250 m spatial resolution MODIS (16-day) NDVI images were extracted for Victoria, over the calendar year January 2008 to January 2009, from Australian mosaics (produced using the methodology described in [56]). To generate the temporal variance in NDVI, a one standard deviation raster was calculated from each annual multi-temporal image pixel-stack.

Elevation (metres), slope (degrees) and aspect (degrees) were derived from a one second (~30 m) smoothed digital elevation model [53]. Climate surfaces were generated using the BIOCLIM component of the ANUCLIM (version 5.1) software package [57], a correlative modelling tool that interpolates climate parameters using spatially explicit digital elevation data and point-based long-term monthly averages of climate variables. A full description of the process can be found in [36,57]. Elevation data raster cells were resampled to 250 m (an appropriate resolution for the distribution of climate stations across the study area) and used as an input to run the BIOCLIM climate model. A subset of the 35 climatic parameters generated by BIOCLIM was selected for inclusion in the model associated with precipitation, temperature, radiation and moisture. BIOCLIM and MODIS NDVI variance surfaces were resampled from 250 m spatial resolution, using the nearest neighbour method, to align with the 30 × 30 m Landsat TM data, elevation layers and textural indices.

4.4. Data Collation

Training and predictor variable data were collated in a GIS database—open-source GRASS Geographic Resources Analysis Support System [42]—and exported into statistics package R [43] for model implementation and analysis, together with training sample raster pixel centroid coordinates. To reduce data redundancy and facilitate interpretation of the model, Pearson correlation coefficients were calculated between all paired combinations of predictor variables. Highly correlated variables ($r^2 > 0.9$, $p < 0.001$) were further examined to calculate biserial correlation coefficients between these predictor variables and a dichotomous forest/non-forest training sample class. Of the highly correlated variable pairs, those with the weaker forest/non-forest relationship were excluded from the model. Table 1 shows the final predictor variables used in the RF model. Variables excluded from the model were the climate layers mean diurnal range, temperature seasonality and annual mean radiation; and textural indices variance (5×5 and 7×7 windows), diversity (3×3 and 7×7 windows) and interspersion (3×3 , 5×5 and 7×7 windows).

Table 1. Random Forests (RF) predictor variables.

Predictor Variable	Units/Data Source	Spatial Resolution (m)
<i>Surface Reflectance</i>		
Landsat TM band 1	0.45–0.52 μm	30
Landsat TM band 2	0.52–0.60 μm	30
Landsat TM band 3	0.63–0.69 μm	30
Landsat TM band 4	0.76–0.90 μm	30
Landsat TM band 5	1.55–1.75 μm	30
Landsat TM band 7	2.08–2.35 μm	30

Table 1. Cont.

Predictor Variable	Units/Data Source	Spatial Resolution (m)
<i>Textural Indices</i>		
Variance (3 × 3)		30
Variance (5 × 5)	Landsat TM NDVI	30
Diversity (3 × 3)		30
<i>Phenological Variability</i>		
NDVI Variance	MODIS NDVI	250
<i>Topography and Climate</i>		
Elevation	SRTM DEM	30
Slope	SRTM DEM	30
Aspect	SRTM DEM	30
Annual Precipitation	mm	250
Annual Temperature Range	°C	250
Annual Mean Temperature	°C	250
Annual Mean Moisture Index	0–1	250

4.5. Random Forest Model

4.5.1. Construction and Evaluation

The *randomForest* package [58] in R [43] was used to build the RF model, for which there are several adjustable implementation parameters. The primary parameters being (i) number of predictor variables randomly sampled as candidates at each decision tree node split (parameter *mtry*); (ii) the number of decision trees (or base classifiers) constructed as part of the classifier ensemble (parameter *ntree*); and (iii) the type of model—classification, regression or unsupervised (parameter *type*). For model construction in this study, the default *mtry* value was used (equal to the square root of the total number of predictor variables). To optimize the number of trees (*ntree*) constructed in the final model, an initial decision tree ensemble was produced with 1,000 trees. Error estimates from the OOB sample showed stabilization of the overall error at 100 trees; therefore, 100 was used for the parameter *ntree* in the final model.

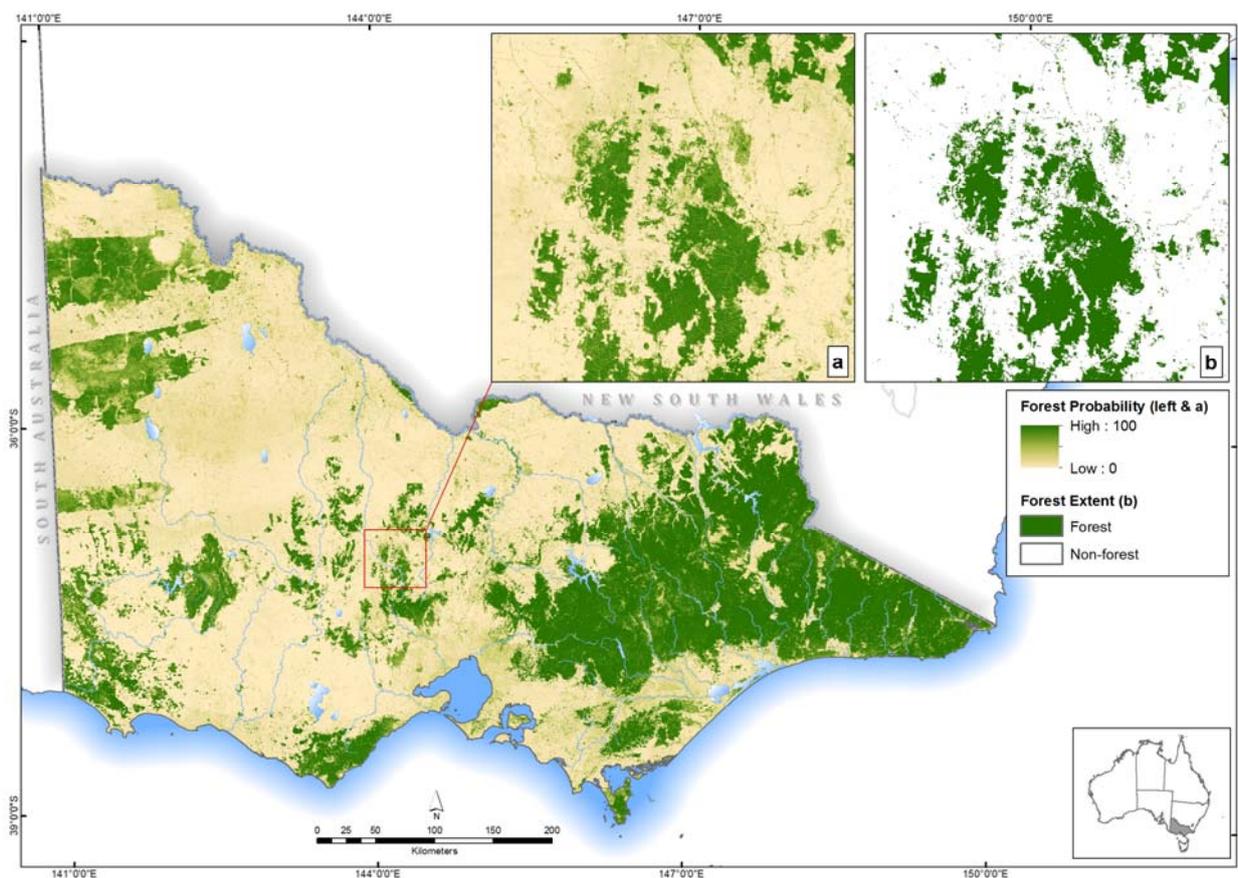
In addition to the RF model OOB test data, for performance evaluation, a 25% subset of training data was randomly sampled, left out of the training dataset (stratified evenly by forest and non-forest classes). The R package *PresenceAbsence* [59] was used to calculate the optimal threshold for converting forest probability (0–100) into a binary forest/non-forest classification, based on maximum Kappa. Kappa, percent correctly classified, user's and producer's accuracy and area under receiver operator curve were calculated to evaluate classification performance. The area under receiver operator curve (ROC) is a measure of a model's ability to discriminate presence (*i.e.*, forest) and absence (*i.e.*, non-forest) [60], calculated from predicted forest probabilities. The ROC is a plot of sensitivity (true positive rate) against specificity (false positive rate). Poor model performance (*i.e.*, where predictive ability is essentially random) returns a near-diagonal ROC plot (true positive rate equal to false positive rate). The area under ROC curve ranges from 0.5 (poor) up to 1. Producer's accuracy (or omission error, one minus producer's accuracy) is the proportion of a land cover class on the ground (*i.e.*, reference) that is correctly classified in the map (prediction). User's accuracy (or commission

error, one minus user's accuracy), is the proportion of a mapped (predicted) class on a map, which matches the corresponding class on the ground (reference). Producer's accuracy measures classification scheme accuracy, while user's accuracy measures the output map generated from the classification [61].

4.5.2. Implementation

The RF model was implemented to predict and map forest probability across the study area. As R holds objects in virtual memory, there are limitations on the resources available for data processing. Therefore, the *RPy* Python package [62] was used, allowing R functionality to be managed within the Python environment outside of R. The study area was divided into two hundred 40 km² tiles, and the RF model was implemented using parallel processing to calculate forest probability across multiple tiles simultaneously, after which the forest probability tiles were mosaicked together into a single forest probability layer.

Figure 3. Implemented Random Forests model forest probability map (a) inset forest probability map (0–100); (b) final forest classification, based on a binary threshold.



Probability values (calculated from the proportion of decision tree votes among all base classifiers in the ensemble) were converted into binary forest and non-forest classes using the probability threshold calculated to maximise the Kappa statistic. To apply the forest definition 0.5 ha minimum mapping unit (MMU) and remove noise from the map, the forest/non-forest classification raster was first re-sampled from 30 m to 28.86 m, so that a 0.5 ha MMU area comprised six whole raster pixels. Horizontally, vertically and diagonally contiguous forest and non-forest cells were grouped together

and attributed a count of the cells within each group. Raster cells within forest cell groups comprising less than six cells (*i.e.*, less than 0.5 ha) were re-labelled as non-forest, and raster cells within non-forest cell groups comprising less than 6 cells were re-labelled as forest. Figure 3 shows the forest probability and final binary forest/non-forest maps.

5. Results and Discussion

5.1. Classification Accuracy

Overall accuracy (percent correctly classified) and Kappa results were high for forest and non-forest prediction using the RF model. Overall accuracy of 96% was achieved, with a Kappa coefficient of 0.91. The threshold value for converting continuous forest probability scores into forest/non-forest classes, optimized to maximize overall Kappa, was 0.5. User's accuracy was marginally higher for the forest class than the non-forest class, indicating a greater tendency for the model to misclassify non-forest land cover as forest, leading to a slight overestimation of forest extent. A comparison of model performance (user's and producer's accuracy) between the test data and the RF OOB accuracy assessment shows marginally lower producer's and user's accuracy for non-forest classification, and user's accuracy in the forest class was returned by the OOB; however, differences between the two accuracy assessment data sources are minor.

The high Kappa coefficient (0.91) for the forest/non-forest classification model is encouraging, and the model accuracy performance is consistent with studies that have successfully discriminated forest from non-forest land cover categories in other natural environments using RF [22,63]. The area under curve (AUC) score (0.91) shows that the RF forest/non-forest classifier has excellent overall model accuracy.

Table 2. Random Forests accuracy assessment. CI, confidence interval; OOB, out-of-bag.

Kappa (CI 95%)	0.914 (0.909–0.919)	
AUC (CI 95%)	0.992 (0.991–0.992)	
Percent Correctly Classified (CI 95%)	95.7 (95.4–95.9)	
	Forest	Non-forest
Kappa maximised binary threshold value	0.5	
Sensitivity	94.42	96.94
Specificity	96.94	94.42
Test (Validation Data)		
Producer's accuracy (omission)	94.42	96.94
User's accuracy (commission)	96.86	94.56
Test OOB		
Producer's accuracy	94.60	96.44
User's accuracy	96.51	94.49

5.2. Variable Importance

Landsat TM band 5 (shortwave infrared) was shown to be the most important variable in predicting forest (Figure 4(a)) based on the calculated mean decrease in accuracy (MDA) score. Band 5 was

considerably more important than the next most important predictor variables—Landsat TM bands 2, 3 and 7, followed by elevation and the four climate surfaces. The high importance of the middle-infrared band 5 (1.55–1.75 μm) in differentiating forest from non-forest at the pixel-level is likely to be associated with its vegetation and soil moisture sensitivity properties. For non-forest classification, based on MDA, elevation was the most important variable in the RF model, followed by bands 2 and 5. The influence of elevation may be associated with less rainfall at lower elevations, but is also very likely to reflect the land use history of the study area, whereby low flat land productive agricultural land has been extensively cleared [64]. Landsat bands 5, 2, 3 and 7 were the most important predictor variables for forest/non-forest differentiation (Figure 4(c)).

Landsat TM band 2 was the most important predictor variable, followed closely by band 5, based on the mean decrease Gini (MDG) measure (calculated for each predictor variable as the cumulative increase in data purity associated with each decision tree node split). Bands 3 and 4 were the next most important variables, followed by NDVI variance and band 7. In comparing the variable importance ranks between the two measures, MODIS NDVI variance was ranked 7 places higher in the MDG measure compared to MDA and band 4 (near-infrared), six places higher. These bands can be considered more important with respect to increasing the purity of training data samples after splitting at decision tree nodes, but less important based on the mean decrease accuracy.

Figure 4. Random Forests predictor variable importance measures. (a) Mean decrease accuracy for forest prediction; (b) mean decrease accuracy for non-forest prediction; (c) mean decrease accuracy for forest and non-forest prediction; and (d) mean decrease Gini for forest and non-forest prediction.

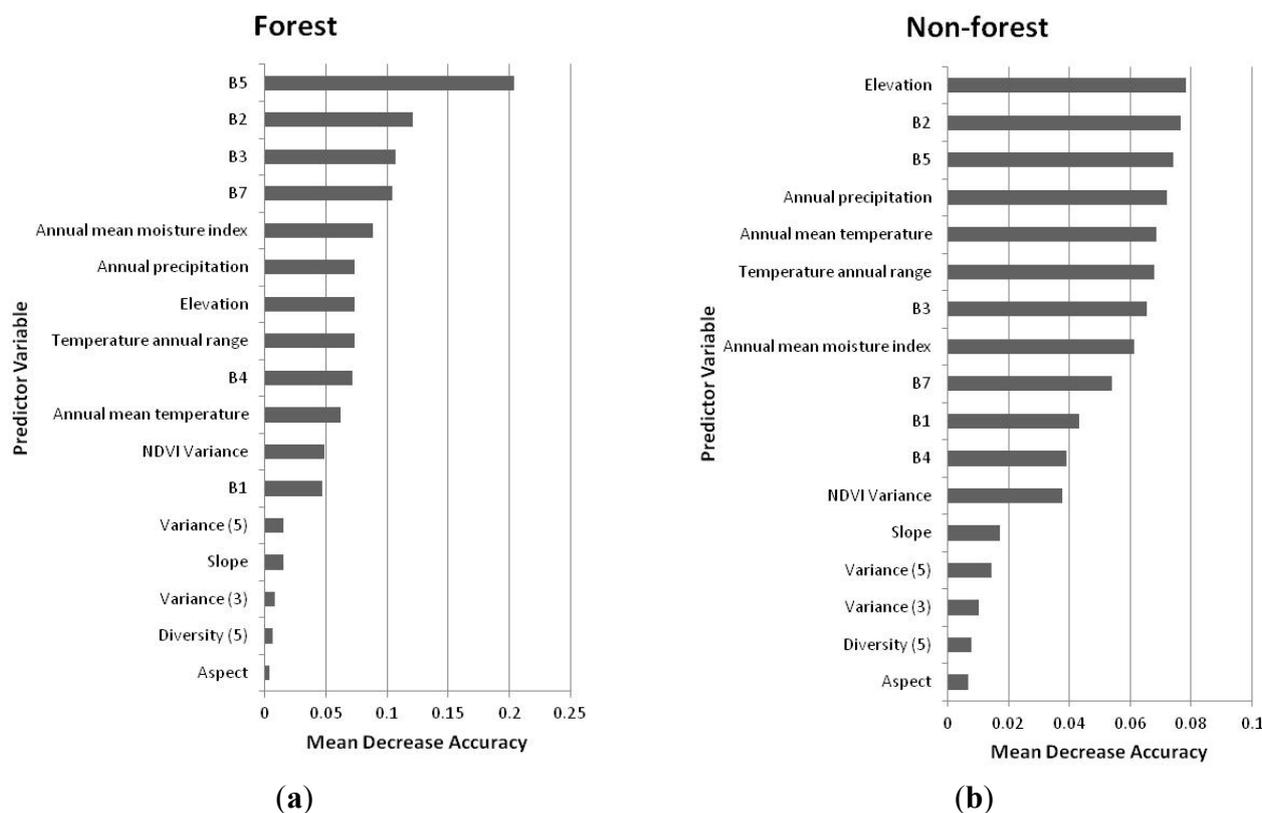
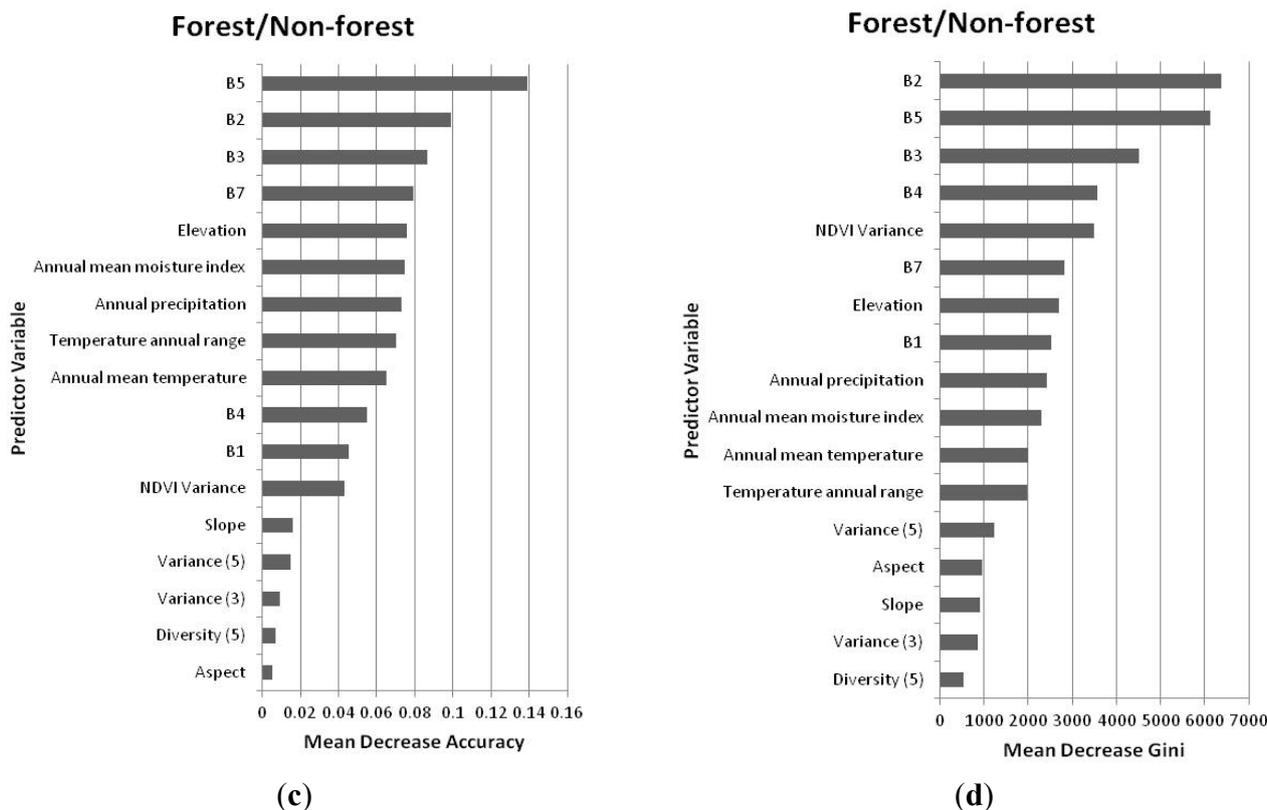


Figure 4. Cont.



The MODIS NDVI variance was included in the model as a means of discriminating seasonally dynamic grasses and understory vegetation from more phenologically ‘stable’ forest canopy reflectance. While results rank this variable as having a reasonably high degree of importance in decision tree node splitting (Gini purity), the low spatial resolution of this layer (250 m) and high spectral heterogeneity within MODIS pixels is likely to be a factor in its lower MDA importance ranking for forest prediction.

Results of this study on application of RF for large area forest classification are encouraging and demonstrate the classifier’s utility in an operational land management agency context. Our results confirm the findings of other studies using RF, that this ensemble classifier can be used to learn complex non-linear relationships. Variable importance measures demonstrate the successful integration of multiple sources of data in predicting forest—remote sensing spectral data and contextual topographic-climate variables.

This study demonstrates the feasibility of using an open-source framework for constructing and evaluating an RF model and its implementation to produce an accurate operational land management agency forest cover map. The framework established successfully integrates freely available spatial data—pre-processed and collated in GRASS—into the R statistical analysis environment. After construction and validation of an RF classifier, the resulting model was implemented in GRASS using an R-GRASS interface package, *spgrass* [44], before finally using GRASS to filter the forest prediction map and apply the minimum mapping unit of the adopted forest definition to the final forest extent spatial product.

In this study, we evaluated the operational performance and utility of the ensemble decision tree classifier, Random Forests (RF), for producing an accurate large area (about 220,000 km²) land

management agency forest map. This study is unique in demonstrating the operational implementation of RF at the regional-scale within an open-source software framework, using GRASS GIS [42] and R [43] statistics software. The framework described, comprising stages of data pre-processing, collation, modelling, evaluation and implementation, contributes to the deployment of affordable programs for collating and processing large volumes of multi-source remote sensing and ancillary GIS data to produce consistent and accurate forest cover maps across complex, noisy and heterogeneous landscapes.

We incorporated Landsat TM and MODIS satellite imagery, textural indices, modelled climate surfaces and topographic layers into an RF model, to accurately predict and map forest across an area comprising millions of hectares of complex and highly diverse forest ecosystems over varying topography, dominated by open canopy sclerophyll forests and woodland. Sample aerial photography land cover maps were used to derive training and test (validation) data. The overall accuracy and Kappa statistics for forest/non-forest classification were 96% and 0.91, respectively. Forest classification achieved a producer's accuracy of 96% and a user's accuracy of 94%. Estimated predictor variable importance measures derived from the Gini Index and out-of-bag (OOB) training data, showed Landsat TM bands 5 and 2 to have the strongest influence in forest/non-forest class-separability.

6. Conclusions

Results show how the RF algorithm can be effectively used to learn the conditional, complex and non-linear relationships between forest vegetation and biophysical factors, to build an accurate forest classifier across highly diverse and dynamic ecosystems. In a land management agency context, the study demonstrates how the RF can be used to address the challenges and operational constraints of land cover classification, including the use of non-parametric and noisy data, its implementation using open-source software, and the integration of multi-source regional scale ancillary spatial data.

While these results are encouraging for the application of RF in an applied natural resource management context, there are several important areas of further research that warrant further investigation. Based on the "Strong Law of Large Numbers", Breiman [1] showed that RF does not over-fit training data as more trees are grown. While results from OOB accuracy and test data support this, the performance of the RF model is based on the important assumption that training data is representative of forest and non-forest classes from across the study area. As proposed by Armston *et al.* [65], in a study investigating the use of RF regression analysis to predict overstory foliage projective cover (FPC) from Landsat TM and ETM imagery, an important next step would be to undertake an independent assessment of the implemented classification model (forest extent map, Figure 4) from sites located away from training data. This would improve understanding of the extent to which spatial autocorrelation between training data samples (*i.e.*, contiguous or closely located pixels) lead to bias, as well as reduced variance and representativeness [66]. In short, how do spatially auto-correlated model training and validation data over-estimate the accuracy and performance of the RF classifier across large heterogeneous landscapes? Other important directions for further research include: (1) the characteristics of RF training data, to better understand how the classifier manages noise and outliers; (2) understanding how different sampling techniques affect classifier performance; and

(3) the implementation of the classifier model on other acquired and calibrated remote sensing image dates and its utility for producing accurate multi-temporal forest extent maps in a monitoring context.

Acknowledgments

The authors would like to acknowledge the support of the Joint Remote Sensing Research Program (notably Neil Flood) for the provision of standardised surface reflectance Landsat TM imagery and support in the implementation of the random forest classifier. We would also like to thank Phil Wilkes for his role in this study.

Conflict of Interest

The authors declare no conflict of interest.

References and Notes

1. McRoberts, R.E. Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sens. Environ.* **2010**, *114*, 1017–1025.
2. Howell, C.I.; Wilson, A.D.; Davey, S.M.; Eddington, M.M. Sustainable forest management reporting in Australia. *Ecol. Indic.* **2008**, *8*, 123–130.
3. Deppe, F. Forest area estimation using sample surveys and Landsat MSS and TM data. *Photogramm. Eng. Remote Sensing* **1998**, *64*, 285–292.
4. Department of Agriculture Fisheries and Forestry. *Australia's Forest at a Glance*; Department of Agriculture Fisheries and Forestry: Canberra, Australia, 2012.
5. Australian Surveying and Land Information Group. *Atlas of Australian Resources (Vol. 6, Vegetation)*; Australian Surveying and Land Information Group: Canberra, Australia, 1990.
6. Jenkins, R.B.; Coops, N.C. Landscape controls on structural variation in Eucalypt vegetation communities: Woronora Plateau, Australia. *Aust. Geogr.* **2011**, *42*, 1–17.
7. Jacobs, M. *Growth Habits of the Eucalypts*; Forestry and Timber Bureau: Canberra, 1955.
8. Behn, G.; McKinnell, F.; Caccetta, P. Mapping forest cover, Kimberley Region of Western Australia. *Australian Forestry* **2001**, *64*, 80–87.
9. Bhandari, S. Monitoring Forest Dynamics using Time Series of Satellite Image Data in Queensland, Australia. PhD Dissertation, The University of Queensland, Brisbane, QLD, Australia, 2011.
10. Jupp, D.L.B.; Walker, J. Detecting Structural and Growth Changes in Woodlands and Forests: The Challenge for Remote Sensing and the Role of Geometric-Optical Modelling. In *The Use of Remote Sensing in the Modeling of Forest Productivity*; Shimoda, H., Gholz, H.L., Nakane, K., Eds.; Springer: Dordrecht, The Netherlands, 1997; pp. 75–108.
11. Montreal Process Implementation Group for Australia. *Australia's State of the Forests Report 2008*; Montreal Process Implementation Group for Australia: Canberra, ACT, Australia, 2008.
12. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.
13. Clerici, N.; Weissteiner, C.J.; Gerard, F. Exploring the use of MODIS NDVI-based phenology indicators for classifying forest general habitat categories. *Remote Sens.* **2012**, *4*, 1781–1803.

14. Main-Knorn, M.; Moisen, G.G.; Healey, S.P.; Keeton, W.S.; Freeman, E.A.; Hostert, P. Evaluating the remote sensing and inventory-based estimation of biomass in the western carpathians. *Remote Sens.* **2011**, *3*, 1427–1446.
15. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm.* **2012**, *67*, 93–104.
16. Austin, M.P.; Meyers, J.a. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *Forest Ecol. Manag.* **1996**, *85*, 95–106.
17. Khalyani, A.H.; Falkowski, M.J.; Mayer, A.L. Classification of Landsat images based on spectral and topographic variables for land-cover change detection in Zagros forests. *Int. J. Remote Sens.* **2012**, *33*, 6956–6974.
18. Cutler, D.R.; Edwards, T.C., Jr.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792.
19. Joy, S.M.; Reich, R.M.; Reynolds, R.T. A non-parametric supervised classification of vegetation types on the Kaibab National Forest using decision trees. *Int. J. Remote Sens.* **2003**, *24*, 1835–1852.
20. Sesnie, S.E.; Gessler, P.E.; Finegan, B.; Thessler, S. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sens. Environ.* **2008**, *112*, 2145–2159.
21. Fahsi, A.; Tsegaye, T.; Tadesse, W.; Coleman, T. Incorporation of digital elevation models with Landsat-TM data to improve land cover classification accuracy. *Forest Ecol. Manag.* **2000**, *128*, 57–64.
22. Gislason, P.; Benediktsson, J.; Sveinsson, J. Random Forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300.
23. Green, G.M.; Sussman, R. Deforestation history of the eastern rainforests of Madagascar from satellite images. *Science* **1990**, *248*, 212–215.
24. Boyd, D.S.; Danson, F.M. Satellite remote sensing of forest resources: Three decades of research development. *Progr. Phys. Geogr.* **2005**, *29*, 1–26.
25. Lu, D. Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. *Int. J. Remote Sens.* **2005**, *26*, 2509–2525.
26. Tucker, C.J.; Townshend, J.R. Strategies for tropical forest deforestation assessment using satellite data. *Int. J. Remote Sens.* **2000**, *21*, 1461–1472.
27. Rogan, J. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. *Remote Sens. Environ.* **2002**, *80*, 143–156.
28. Maselli, F. Use of MODIS NDVI data to improve forest-area estimation. *Int. J. Remote Sens.* **2011**, *32*, 6379–6393.
29. Wulder, M.a; White, J.C.; Gillis, M.D.; Walsworth, N.; Hansen, M.C.; Potapov, P. Multiscale satellite and spatial information and analysis framework in support of a large-area forest monitoring and inventory update. *Environ. Monit. Assess.* **2010**, *170*, 417–433.
30. Culbert, P.; Pidgeon, A.; St-Louis, V. The impact of phenological variation on texture measures of remotely sensed imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2009**, *2*, 299–309.
31. Coburn, C. a.; Roberts, a. C. B. A multiscale texture analysis procedure for improved forest stand classification. *Int. J. Remote Sens.* **2004**, *25*, 4287–4308.

32. Eckert, S. Improved forest biomass and carbon estimations using texture measures from worldview-2 satellite data. *Remote Sens.* **2012**, *4*, 810–829.
33. Kayitakire, F.; Hamel, C.; Defourny, P. Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery. *Remote Sens. Environ.* **2006**, *102*, 390–401.
34. Rodríguez-Galiano, V.F.; Abarca-Hernández, F.; Ghimire, B.; Chica-Olmo, M.; Atkinson, P.M.; Jeganathan, C. Incorporating Spatial Variability Measures in Land-cover Classification using Random Forest. *Procedia Environ. Sci.* **2011**, *3*, 44–49.
35. Guisan, A.; Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecol. Model.* **2000**, *135*, 147–186.
36. Beaumont, L.; Hughes, L.; Poulsen, M. Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. *Ecol. Model.* **2005**, *186*, 250–269.
37. Franklin, J. Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. *Progr. Phys. Geogr.* **1995**, *19*, 474–499.
38. Random Forest. Available online: http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (accessed February 9, 2012).
39. Calle, M.L.; Urrea, V. Letter to the editor: Stability of Random Forest importance measures. *Briefings Bioinf.* **2011**, *12*, 86–9.
40. The GNUManifesto. Available online: <http://www.gnu.org/gnu/manifesto.html> (accessed on 10 February 2011).
41. Rocchini, D.; Delucchi, L.; Bacaro, G.; Cavallini, P.; Feilhauer, H.; Foody, G.M.; He, K.S.; Nagendra, H.; Porta, C.; Ricotta, C.; *et al.* Calculating landscape diversity with information-theory based indices: A GRASS GIS solution. *Ecol. Inform.* **2012**, in press.
42. GRASS Development Team. Geographic Resources Analysis Support System (GRASS) Software; Version 6.4; Open Source Geospatial Foundation Project; 2012. Available online: <http://grass.osgeo.org> (accessed on 17 April 2013).
43. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria. 2011. Available online: <http://www.R-project.org> (accessed on 17 April 2013).
44. Bivand, R. Using the R-GRASS Interface: Current Status. *OSGeo Journal* **2007**, *1*, 36–38.
45. The Python Language Reference. Available online: <http://docs.python.org/release/3.2/reference/index.html> (accessed on 27 May 2013).
46. Viridans Ecosystems and Vegetation. Available online: <http://www.viridans.com/ECOVEG/> (accessed on 27 May 2013).
47. Department of Sustainability and Environment Victorian Forest Monitoring Program. Available online: <http://www.dse.vic.gov.au/forests/managing-our-forests/forest-sustainability/victorian-forest-monitoring-program> (accessed on 10 December 2012).
48. Mellor, A.; Haywood, A. Remote Sensing Victoria's Public Land Forests—A Two Tiered Synoptic Approach. In Proceedings of the 15th Australian Remote Sensing and Photogrammetry Conference; Alice Springs, Australia, 13 September 2010.
49. National Forest Inventory. *Australia's State of the Forests Report 2003*; Bureau of Rural Sciences: Canberra, ACT, Australia, 2003.

50. Food and Agriculture Organization of the United Nations. *Global Forest Resources Assessment 2000*; FAO: Rome, Italy, 2001; p. 479.
51. Farmer, E.; Jones, S.; Clarke, C.; Buxton, L.; Soto-Berelov, M.; Page, S.; Mellor, A.; Haywood, A. Creating A Large Area Landcover Dataset For Public Land Monitoring And Reporting. In *Progress in Geospatial Science Research*; Arrowsmith, C., Bellman, C., Cartwright, W., Jones, S., Shortis, M., Eds.; Publishing Solutions: Melbourne, VIC, Australia, 2013; pp. 85–98.
52. Earth Explorer. Available online: <http://earthexplorer.usgs.gov> (accessed on 27 May 2013).
53. CSIRO One-second SRTM digital elevation model. Available online: <http://www.csiro.au/Outcomes/Water/Water-information-systems/One-second-SRTM-Digital-Elevation-Model.aspx> (accessed on 27 May 2013).
54. Flood, N.; Danaher, T.; Gill, T.; Gillingham, S. An operational scheme for deriving standardised surface reflectance from Landsat TM/ETM+ and SPOT HRG imagery for Eastern Australia. *Remote Sens.* **2013**, *5*, 83–109.
55. Haralich, R.M. Statistical and structural approach to texture. *Proc. IEEE* **1979**, *67*, 786–804.
56. Paget, M.J.; King, E.A. *MODIS Land Data Sets for the Australian Region*; CSIRO Marine and Atmospheric Research: Canberra, ACT, Australia, 2008.
57. Houlder, D.; Hutchinson, M.; Nix, H.; McMahon, J. *ANUCLIM*; Version 5.1; Centre for Resource and Environmental Studies: Canberra, ACT, Australia, 2001.
58. Liaw, A.; Wiener, M. Classification and regression by RandomForest. *R News* **2002**, *2*, 18–22.
59. Freeman, E.A.; Moisen, G. PresenceAbsence: An R package for Presence-Absence Model analysis. *J. Stat. Softw.* **2008**, *23*, 1–31.
60. Pearce, J.; Ferrier, S. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* **2000**, *133*, 225–245.
61. Shao, G.; Wu, J. On the accuracy of landscape pattern analysis using remote sensing data. *Landscape Ecol.* **2008**, *23*, 505–511.
62. RPy Python interface to the R Programming Language. Available online: <http://rpy.sourceforge.net> (accessed on 12 May 2012).
63. Chan, J.C.-W.; Paelinckx, D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens. Environ.* **2008**, *112*, 2999–3011.
64. Woodgate, P.; Black, P. *Forest Cover Changes in Victoria 1869–1987*; Remote Sensing Group, Lands and Forests Division, Dept. of Conservation, Forests and Lands: Melbourne, VIC, Australia, 1988; p. 31.
65. Armston, J. D. Prediction and validation of foliage projective cover from Landsat-5 TM and Landsat-7 ETM+ imagery. *J. Appl. Remote Sens.* **2009**, *3*, 033540.
66. Chen, D.; Stow, D. The effect of training strategies on supervised classification at different spatial resolutions. *Photogramm. Eng. Remote Sensing* **2002**, *68*, 1155–1161.