

## Article

# Mapping Physiognomic Types of Indigenous Forest using Space-Borne SAR, Optical Imagery and Air-borne LiDAR

John R. Dymond <sup>1,\*</sup>, Jan Zörner <sup>2</sup> , James D. Shepherd <sup>1</sup>, Susan K. Wiser <sup>3</sup>, David Pairman <sup>3</sup> and Marmar Sabetizade <sup>4</sup>

<sup>1</sup> Manaaki Whenua—Landcare Research, Palmerston North 4410, New Zealand

<sup>2</sup> Manaaki Whenua—Landcare Research, Wellington 6143, New Zealand

<sup>3</sup> Manaaki Whenua—Landcare Research, Lincoln 7608, New Zealand

<sup>4</sup> Department of Soil Science, University of Tehran, Tehran 1417, Iran

\* Correspondence: dymondj@landcareresearch.co.nz

Received: 15 July 2019; Accepted: 12 August 2019; Published: 15 August 2019



**Abstract:** Indigenous forests cover 24% of New Zealand and provide valuable ecosystem services. However, a national map of forest types, that is, physiognomic types, which would benefit conservation management, does not currently exist at an appropriate level of detail. While traditional forest classification approaches from remote sensing data are based on spectral information alone, the joint use of space-based optical imagery and structural information from synthetic aperture radar (SAR) and canopy metrics from air-borne Light Detection and Ranging (LiDAR) facilitates more detailed and accurate classifications of forest structure. We present a support vector machine (SVM) classification using data from the European Space Agency (ESA) Sentinel-1 and 2 missions, Advanced Land Orbiting Satellite (ALOS) PALSAR, and airborne LiDAR to produce a regional map of physiognomic types of indigenous forest. A five-fold cross-validation (repeated 100 times) of ground data showed that the highest classification accuracy of 80.5% is achieved for bands 2, 3, 4, 8, 11, and 12 from Sentinel-2, the ratio of bands VH (vertical transmit and horizontal receive) and VV (vertical transmit and vertical receive) from Sentinel-1, and mean canopy height and 97th percentile canopy height from LiDAR. The classification based on optical bands alone was 72.7% accurate and the addition of structural metrics from SAR and LiDAR increased accuracy by 7.4%. The classification accuracy is sufficient for many management applications for indigenous forest, including biodiversity management, carbon inventory, pest control, ungulate management, and disease management.

**Keywords:** forest types; forest mapping; Sentinel-2; SAR; LiDAR; canopy metrics

## 1. Introduction

Indigenous forests cover 24% of New Zealand, primarily in mountainous and hilly terrain, representing a 70% reduction from the pre-human state (ca. 800 years ago) due to fire, forest clearance, and logging [1]. These forests provide highly valued ecosystem services, including recreation, sense of belonging, tourism, and important provisioning and regulating services, such as wild foods, fresh water, and regulation of both climate and soil erosion [2]. They contain much of the unique biodiversity in New Zealand where 78% of native plants are endemic and 91% of terrestrial Animalia are endemic [3]. The evergreen forests are warm and temperate in the far north and progress through to cool and temperate in the south [4]. Unfortunately, many plants and animals introduced to New Zealand have become pests in indigenous forests. Some browsing animal pests (such as Australian brushtail possums, pigs, and deer) impact forest plants directly, whereas animal predators (such as cats, rats and stoats) impact the indigenous fauna in the forests. Recently, two tree diseases—kauri dieback

(*Phytophthora agathidicida*) and South American myrtle rust (*Austropuccinia psidii*)—have raised concern in New Zealand; these also endanger some of the tall conifer and broadleaved trees (*Agathis australis*, *Metrosideros robusta*, *Metrosideros umbellata*).

Maps of forest types can be used to address a wide range of ecological questions relating to the sustainability of ecosystem services and conservation management. An example is large-scale pest control (costing > \$20 million), in New Zealand beech forests after beech masting years [5] when excessive seed production stimulates pest populations. Detailed maps of beech forests are required for planning control operations. Both the New Zealand Forest Class Maps (compiled at a scale of 1:250,000 from black and white aerial photographs, dating from 1948 to 1955; available online [6]) and the Vegetative Cover Map of New Zealand (compiled at a scale of 1:1,000,000 [7]) provide national coverage of broad forest types, but are now 50+ years out of date, and the scale is not fine enough for many desired applications. Wiser et al. [8,9] used data from > 13,000 ground-based plots to classify New Zealand woody vegetation into forest types (at two levels of hierarchy termed alliances and associations). They then grouped these into more broadly defined physiognomic types as Beech, Beech-broadleaved, Beech-broadleaved-podocarp, Broadleaved-podocarp, Podocarp, and other forests. This classification has yet to be spatially extrapolated to produce a national map at any level of the hierarchy at an appropriate scale (~1:50,000). EcoSat Forests [10], which is a national map of indigenous forest classes in New Zealand, is published at 1:750,000 and the Land Cover Database (LCDB) [11] for New Zealand has only one indigenous forest class. Spatial extrapolation of plot-based classification of vegetation could potentially be achieved by linking the vegetation types to spectral responses from multispectral satellite imagery [11]. To date, a primary barrier to this has been that spectral variation of evergreen forests in optical satellite imagery is generally greater within physiognomic types than between them [12].

An ever-growing fleet of earth observation satellites is acquiring data with improved spectral, radiometric, and spatial resolutions, and with shorter revisit times. Satellite missions aiming at high (~1 m) spatial resolution, such as WorldView [13], GeoEye [13], RapidEye [13], and SkySat [14], minimize the instantaneous field of view (IFOV) of the sensor at the cost of a small swath width and thus are limited in the spatiotemporal coverage for large areas, i.e., entire countries. Novel missions such as PlanetScope [15] fly groups of microsatellites with high spatial resolution (~5 m) in constellation and are thus able to capture the entire land surface of the Earth every day. However, these sensors are currently limited to a small set of spectral bands not suitable for fine vegetation discrimination and do not provide a long-term record of data. Land-monitoring satellite sensors with medium spatial resolution (10–100m) have been in operation since the early 1970s (Landsat [16]) and sample the Earth's surface globally in a few days. The European Space Agency (ESA) launched the Sentinel-2 mission in 2015 to continue global coverage of the Earth provided by Landsat and SPOT [17] satellites. The Sentinel-2 mission is of particular interest because of its wide swath (290 km), good revisit time (5 days, with two satellites), high and medium spatial resolution (10, 20, and 60 m), and 13 spectral bands specifically chosen to capture surface and vegetation properties [18].

In contrast to optical satellite imagery, synthetic aperture radar (SAR) can provide information about the near-surface structure of sample objects. For analysis of top-of-canopy structure of forests we use data from ESA's Sentinel-1 A and B platforms, both of which carry C-band synthetic-aperture radar instruments with a 1-decibel (dB) radiometric accuracy and a central frequency at 5.405 GHz [19]. The relatively short wavelength of about 5.5 cm facilitates high sensitivity to the top-most canopy layer. Both satellite platforms are flown in a sun-synchronous polar orbit at 693 km altitude and are jointly able to capture the same ground location with a 6-day repeat cycle. Main measurement mode over land areas is the Interferometric Wide-Swath Mode with 250-km swath, 5 × 20 m (range × azimuth) spatial resolution and burst synchronisation for interferometry. To capture structural information within the canopy lower frequency radar is necessary. We use data from Phased-Array L-band SAR (PALSAR) carried onboard the Advanced Land Orbiting Satellite (ALOS) which was operational from 2006 to

2011 with 24-cm wavelength and acquiring polarimetric data (horizontal and vertical transmit and receive) [20].

Another remote sensing technique able to measure canopy structure is airborne Light Detection And Ranging (LiDAR), also referred to as Airborne Laser Scanning (ALS). In ALS, laser pulses are emitted downwards to the ground and reflected by canopy features back to the sensor, thereby capturing vertical information on canopy structure [21]. In 2013 and 2014, the Wellington Regional Government GIS Group conducted an extensive LiDAR survey with a minimum point density of 1.3 points/m<sup>2</sup> and a vertical accuracy of  $\pm 0.15$  m over the Wellington region (812,000 ha). While their main intention was to build new, large-scale topographic maps of the region [22], the survey also provides an unprecedented opportunity to study forest composition and individual tree parameters on a large scale.

In addition to imagery quality and fitness for purpose [23], the choice of classification method is also important. Nonparametric methods are proving to be effective for classifying forest types [24], especially Random Forests and Support Vector Machine (SVM) classification, due to their ability to synthesize classification functions on discrete and continuous data sets, and to deal with unbalanced data, and also to their insensitivity to noise or overtraining [25,26]. These properties enable the combination of spectral data with other types of data to improve the accuracy of classifying forest types. However, research into combining satellite optical data with other remotely sensed data relating to vegetation structure, such as LiDAR and SAR, is incipient [27,28]. Classification of objects or segments rather than pixels, termed GEOBIA (Geographic-Object-Based Image Analysis), has improved classification accuracy in forest applications [29–33]. Temporal analysis has also proved useful [12,34–37].

In this paper, we investigate how the classification of forest physiognomic types using Sentinel-2 may be improved with the addition of other imagery relating to vegetation structure. We consider radar backscatter from Sentinel-1 and PALSAR (Phased Array L-band Synthetic Aperture Radar), and canopy height metrics (mean and 97th percentile) from airborne LiDAR. These data are available for the whole of the Wellington region, New Zealand—the study area. Fusion of the data is made possible by calculating mean values within segments of typical size 5 ha, that is, using a GEOBIA approach. We use SVM classification because of its ability to deal with unbalanced data and superior insensitivity to noise and overtraining [25,26]. With SVM, we explore the best combinations of Sentinel-1, Sentinel-2, PALSAR, and CHM (canopy height model) data by five-fold cross-validation. Using the best performing classifier, we produce a map of forest physiognomic types in the Wellington region along with summary statistics.

## 2. Study Area and Forest Physiognomic Types

The Wellington region, in the lower part of the North Island (Figure 1), New Zealand, has 23% of its 812,000 ha of land, covered in indigenous forest. In pre-Māori times, nearly all the Wellington region was covered by indigenous forest, but after their arrival (ca. 1000 A.D.), a significant proportion was burned for shifting agriculture, and after the arrival of Europeans (ca. 1840 A.D.) an even larger proportion was burned and cleared for pastoral agriculture. Most of the indigenous forest is now confined to the protected mountainous areas of the Tararua, Rimutaka, and Aorangi ranges. Many remnants of indigenous forest are spread throughout the rest of the region, and there are large blocks of shrubland reverting to indigenous forest on the east coast. The climate of the Wellington region is mild and wet: annual rainfall varies between 800 mm in the eastern Wairarapa and 7000 mm at the tops of the Tararua range (ca. 1500 m), and mean monthly temperatures at Wellington city range between 8° C in July and 16° C in January.

The forests of the region are dominated by various mixtures of species from three groups: conifers, all from the Podocarpaceae family; broad-leaved evergreen species from a wide range of families; and Southern beeches (Nothofagaceae). Forests dominated by broad-leaved trees with scattered emergent podocarps (Broadleaved-podocarp forest) occur on most lowland sites (podocarp abundance has

been reduced by historical selective logging in many areas). Broadleaved-podocarp forest is usually luxuriant with small trees, shrubs, and ferns forming the understoreys with abundant mosses, lichens, lianes, and epiphytes. There is a shift to beech dominance at higher elevations, or on sites with dry climates or infertile soils. Beech forest tends to be dominated by just one or more of the southern-beech species and usually has lower diversity of associated plant species. In this paper, we focus on distinguishing forests based on the relative proportions of tree species of differing physiognomy in the canopy. We focus on the five combinations that occur in the Wellington region: Broadleaved-podocarp forest; Beech-broadleaved-podocarp forest; Beech-broadleaved forest; Broadleaved forest; and Beech forest. Podocarp forest is uncommon in the Wellington region, but common in other regions.

### 3. Methods

#### 3.1. Sentinel-2 imagery

We obtained three Sentinel-2 images of the Wellington region with reasonably high sun angle dated 22 November 2017, 22 November 2016, and 28 December 2015. The images were approximately 95% cloud-free and together gave a 99.5% cloud free coverage of the Wellington region. Sentinel-2 imagery has 13 bands, from which we use 7 bands important for classifying vegetation: blue (band 2) ( $490 \pm 50$  nm); green (band 3) ( $560 \pm 25$  nm); red (band 4) ( $665 \pm 20$  nm); red edge (band 5) ( $705 \pm 10$  nm); near infrared (band 8) ( $835 \pm 65$ ); short-wave infrared (band 11-SWIR) ( $1610 \pm 70$  nm); and short-wave infrared (band 12-SWIR) ( $2185 \pm 120$  nm). The visible and near infrared bands have 10-m pixels. The red edge and short-wave infrared bands have 20-m pixels that were pan-sharpened with the visible bands to 10-m pixels [38].

#### 3.2. Processing of Sentinel-2 Imagery to Standardised Reflectance

The radiance (i.e., apparent brightness) of vegetation as seen by a satellite is the product of irradiance and the vegetation reflectance, attenuated through the atmosphere. Both the irradiance and reflectance are influenced by the position of the sun relative to the slope normal of the vegetation surface. As reflectance is also influenced by the viewing direction of the satellite relative to the slope normal, in a satellite image of hilly or mountainous terrain, the apparent radiance of vegetation is strongly influenced by topography. For use in vegetation classification it is desirable for a satellite image to represent only the properties of the vegetation, or other reflecting surfaces. To achieve this, it is necessary to process the satellite image to standardised spectral reflectance, which is the spectral reflectance the vegetation would have on flat terrain with the sun and satellite in standard positions (nadir view for the satellite, and a solar elevation of 50 degrees). We processed the satellite imagery to standardised reflectance using the method of Dymond and Shepherd [38], which requires physical modelling of radiation from the sun and sky through the atmosphere, reflection of the light from the vegetation canopy, and the transmission of the reflected light through the atmosphere to the satellite sensor.

#### 3.3. Segmentation of Sentinel-2 imagery

A cloud-free mosaic of the Wellington region (99.5% cloud free) was produced by extraction of highest priority cloud-free pixels from the Sentinel-2 images (Section 3.1), with highest priority given to dates nearest the middle of November 2017. The mosaic was then segmented using an iterative elimination process. An initial segmentation was produced from k-means clustering ( $k = 60$ ), with small clumps being eliminated iteratively, smallest first, until a minimum mapping unit was achieved. The simple method produces results comparable with other well-known algorithms [39] but has the advantages of scalability to large datasets and a target minimum mapping unit. The method can retain smaller segments than the minimum mapping unit where they are significantly spectrally distinct from neighbours, such as small water bodies. The algorithm has been made available in the open source

software library RSGISLib [40]. The segments are used as a framework for estimating robust means of Sentinel-1, Sentinel-2, and PALSAR imagery, and of canopy height from LiDAR data.

### 3.4. Processing of Sentinel-1 SAR imagery

We use the Sentinel-1 Level-1 Ground Range Detected (GRD) product consisting of focused SAR data that has been detected, multi-looked and projected to ground range. Individual orbit slices are downloaded for the study region and stored on the NeSI high-performance computer (HPC) facility [41]. A pre-processing workflow has been developed to prepare the uncalibrated SAR data for use in this study. The following steps are performed using routines built into the ESA Sentinel Application Platform (SNAP) version 6 and run on the HPC in parallel for individual orbits: (1) thermal noise removal; (2) satellite orbit correction; (3) radiometric calibration; (4) mosaic of individual orbit slices; (5) speckle filtering of radar noise using a single-product Lee-filter (window size:  $7 \times 7$ , sigma: 0.9, target window size:  $3 \times 3$ ); (6) radiometric terrain flattening using NASA's SRTM elevation model; (7) geometric terrain correction and re-projection to NZTM (New Zealand Transverse Mercator) with  $10 \times 10$  m pixels; and (8) conversion to decibel (dB) units and clipping off noise along the near and far range edges (using a custom polygon boundary for each orbit). For radiometric and geometric terrain correction NASA's SRTM 1 Sec product [42] is used.

### 3.5. Processing of PALSAR Imagery

The PALSAR Global Mosaic [24,43] is a high-resolution slope-corrected and ortho-rectified L-band gamma-naught map. The data used for this study were acquired at 10 m resolution in a fine-beam dual-polarization mode (HH and HV) (H=horizontal and V=vertical) for the year 2007. The data were resampled to the NZTM map projection, mosaicked, and calibrated to backscatter power using published calibration factors [32]:

$$\gamma^0 = 10^{2 \log_{10} DN - 8.3} \quad (1)$$

where  $DN$  is the 16-bit unsigned value from the ALOS-PALSAR product. Backscatter was then converted to decibels.

### 3.6. Production of Canopy Height Model

A LiDAR survey was conducted over the entire Wellington region [22], primarily in early 2013 but with some additional aircraft flights later in 2013 and 2014, depending on weather. The LiDAR scanner was an Optech Airborne Laser Terrain Mapper ALTM 3100EA flown at a nominal height of 1000 m above the ground. Target point density was 1.7 points per square meter (ppsm) with 50% swath overlap to ensure the minimum specification of 1.3 ppsm and a vertical accuracy of  $\pm 0.15$  m. While these were the minimum requirements, the actual point density of the dataset ended up higher, ranging between 4 and 6 ppsm on average and  $> 6$  ppsm in regions with overlapping flight paths. Coincident with the LiDAR survey was an aerial photographic survey with blue, green, red and near infrared bands, which was ortho-rectified to 30 cm pixels.

The 1261 LiDAR flight lines of point cloud data were merged, tiled and further processed using the open-source LiDAR processing library, Sorted Pulse Data Library (SPDLib) [44] as described in [45]. A canopy height model (CHM) at  $1 \times 1$  m pixel resolution was generated with outliers being removed by applying a  $5 \times 5$  m Gaussian median filter and screening out pixels below 0.5 m (undesired low vegetation) and above 60 m (erroneous high trees).

### 3.7. Ground Data

Data from 580 vegetation plots in the Wellington region having tree species observations were extracted from the National Vegetation Survey (NVS) database [46]. Where plots are permanent, only the most recent measurement was taken. These spanned the years 1962 to 2017, with 87% of the plots being measured since 1980. Plots subject to experimental treatments (e.g., exclosures) were excluded.



Plots are of size  $20 \times 20$  m and contain measurements of diameters at 1.35 m for all trees of diameter over 20 mm. On 276 of the plots, relevé data (species abundance in cover classes within height tiers) are available [47]. These data were summarised by listing (i) the five most abundant tree species by basal area and cover; (ii) relative basal area and cover of trees classed as podocarps, beech and broadleaved species; and (iii) the vegetation alliance to which the plot belonged. Summaries informed assignment of the plot to a physiognomic type. A polygon-shapefile of physiognomic forest type was generated from this summary data by overlaying visually, using ArcGIS, the plot location on the orthophotography (i.e., the natural colour and colour infrared images acquired together with the LiDAR measurements) and on a colour rendition, according to the height of the canopy height model. The plots typically are located in sets, usually along altitudinal transects located within catchments using a two-stage random/systematic sampling scheme. It was possible to determine the boundaries of the physiognomic type associated with these sets of plots from the visually apparent relationships with the canopy height model and orthophotography. Where the physiognomic type derived from the plot data was inconsistent with the canopy height model and orthophotography (which could arise owing to successional change since the plot was measured or inaccuracies in location), the physiognomic type was inferred from the imagery. We drew polygons of homogeneous physiognomic types of typical size 5–10 ha. These types (hereafter termed ‘ground data’ were assigned to the segments (i.e., the raster attribute table of the segment image) derived in Section 3.3 if a segment was wholly or primarily contained (>90%) in a polygon.

### 3.8. Support Vector Machine Classification

The mean standardised reflectance of each Sentinel-2 band was calculated for each segment and stored in the raster attribute table of the segment image. Likewise, the mean backscatter in each Sentinel-1 band and in each PALSAR band was calculated for each segment and stored in the raster attribute table, along with the mean and the 97th percentile (below which 97% of heights are below) of the canopy height model. Together with the ground data (449 segments in total), these data formed the training and test data. We explored the best combinations of Sentinel-1, Sentinel-2, PALSAR, and CHM data by performing five-fold cross-validation, where the ground data is randomly partitioned into 5 equal sized subsamples and each subsample used to test the remaining 4 subsamples assigned to training. The five-fold cross-validation was repeated 100 times with different random selections of subsamples to get an average overall classification accuracy (which we seek to maximise). For the SVM classification, the R-package e1071 is used as interface to the SVM implementation in the LIBSVM software [48]. A linear kernel and default parameters for cost (= 1) and gamma (= 0.1) were used.

## 4. Results

Table 1 shows all the optical bands, radar bands, and CHM metrics that make a significant contribution to cross-validation accuracy when included in a SVM classification with a linear kernel (significant meaning greater than 1% contribution when combined with other like and significant data, that is, with other optical, with other canopy height, or with other radar data). When all significant optical and radar bands and CHM metrics are included, the accuracy of the SVM classification is 80.05%. The relative importance of each band and metric is shown in the difference between using all bands and metrics and with leaving one out. The most important band/metric is CHM97, the 97th percentile of canopy heights in a segment, with a 4.27% contribution to accuracy. The most important optical bands are B02, the blue band, with a 1.85% contribution to accuracy, and B12, the 2nd short-wave infrared band, with a 1.85% contribution to accuracy also. The most important radar band is the Sentinel-1 VH/VV with a 1.63% contribution to accuracy.

When all the bands and metrics are combined, the red edge band, B05, and PALSAR HH become less significant and make a small contribution to the final accuracy, with B05 making a negative contribution. Table 2 repeats the analysis of Table 1 but omits B05 and PALSAR HH. When only these bands and metrics are combined, the accuracy of the SVM classification increases to 80.48% from

80.05%. The most important band/metric is still CHM97, with an increased 5.19% contribution to accuracy. The most important optical band is now solely B02, with an increased 2.15% contribution to accuracy. The most important radar band is Sentinel-1 VH/VV with an increased 1.80% contribution to accuracy.

Table 3 shows the cross-validation accuracy when using all the significant optical bands, then adding the significant CHM metrics to the significant optical bands, and then adding the significant radar bands to the optical and CHM metrics. The optical bands alone achieve 72.67% accuracy. With the addition of the CHM metrics this is increased to 78.48%, which is a significant increase of 5.81%. With the addition of the significant radar bands to the optical and CHM metrics the accuracy is increased to 80.05%, another significant increase of 1.57%.

**Table 1.** Five-fold cross-validation accuracy (100 repeats) of predicted forest physiognomic types from support vector machine classification by leaving out one from all the significant bands (i) Sentinel-2 spectral bands (2,3,4,5,8,11,12), (ii) mean and 97th percentile of canopy height model, and (iii) Sentinel-1 VH/VV and PALSAR HH bands. Grey shade shows which bands were used.

Sentinel-2 band 2	
Sentinel-2 band 3	
Sentinel-2 band 4	
Sentinel-2 band 5	
Sentinel-2 band 8	
Sentinel-2 band 11	
Sentinel-2 band 12	
Mean of CHM	
97 <sup>th</sup> perc. of CHM	
Sentinel-1 VH/VV	
PALSAR HH	
Accuracy (%)	80.05 80.03 78.42 75.78 78.96 78.20 79.65 79.45 80.44 79.61 79.77 78.20

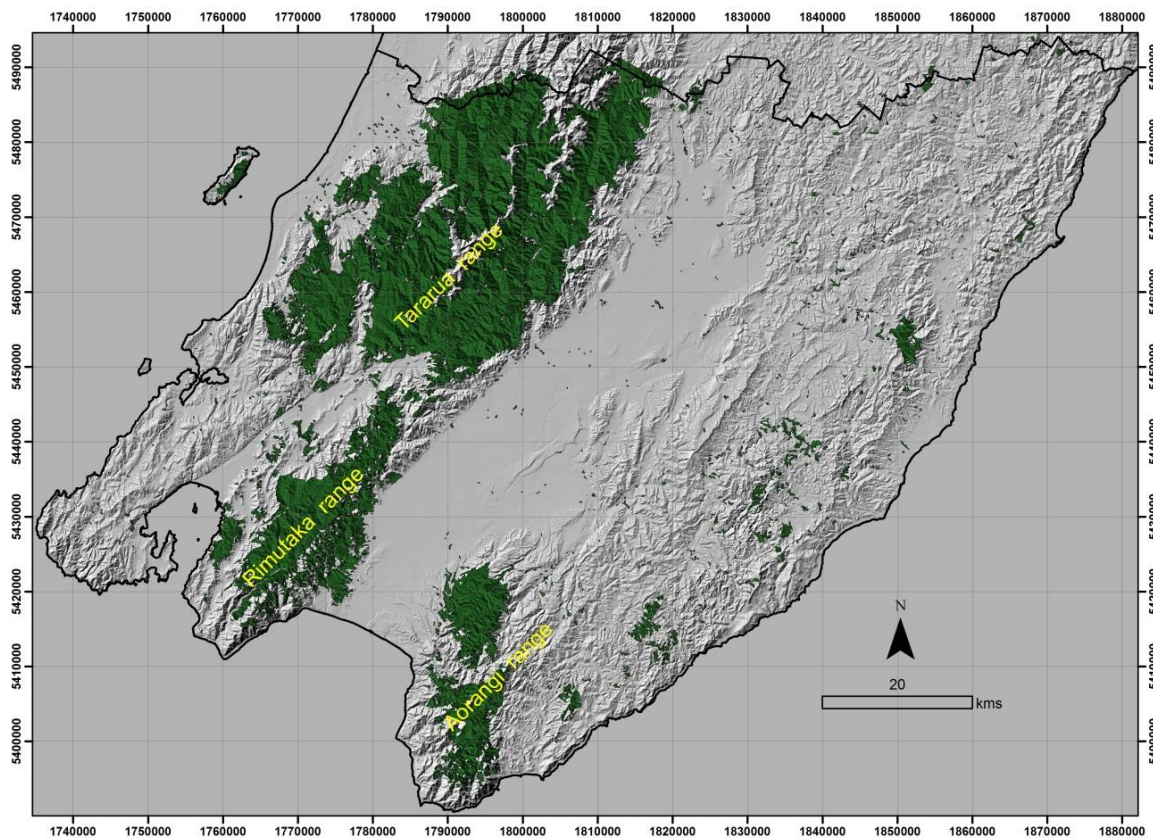
**Table 2.** Five-fold cross-validation accuracy (100 repeats) of predicted forest physiognomic types from support vector machine classification by leaving out one from the best combination of bands, that is from (i) Sentinel-2 spectral bands (2,3,4,8,11,12), (ii) mean and 97th percentile of canopy height model, and (iii) Sentinel-1 VH/VV bands. Grey shade shows which bands were used.

Sentinel-2 band 2										
Sentinel-2 band 3										
Sentinel-2 band 4										
Sentinel-2 band 8										
Sentinel-2 band 11										
Sentinel-2 band 12										
Mean of CHM										
97 <sup>th</sup> perc. of CHM										
Sentinel-1 VH/VV										
Accuracy (%)	80.48	78.68	75.29	78.57	78.68	79.89	79.87	79.90	80.55	78.33

The SVM classifier established above was trained on all locations for which ground data were available and then applied to all the image segments of indigenous forest in the Wellington region (according to the Land Cover Database [11]) to produce a regional map of forest physiognomic types. Figure 2 shows a subset of this map in a transect across the Tararua ranges. The spatial distribution of physiognomic types is consistent with common observations. Broadleaved-podocarp forest is often in valley floors and neighbouring slopes. Progression upwards in elevation corresponds with a transition to Beech-broadleaved-podocarp forest and then to Beech-broadleaved forest. Near the tree line, the forest is usually pure Beech. On the western flanks of the Tararua ranges, near the coast, Broadleaved forest predominates.

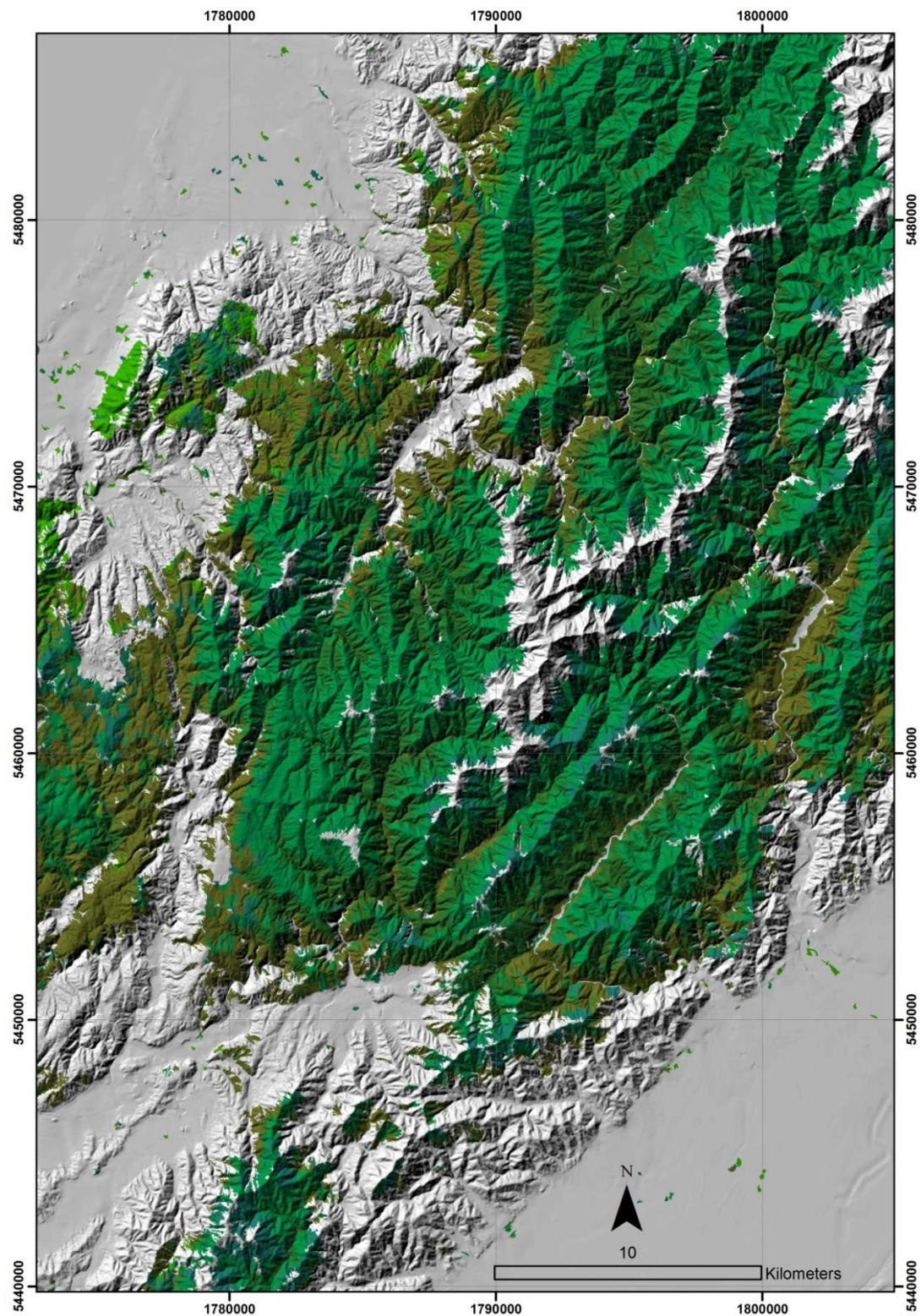
**Table 3.** Five-fold cross-validation accuracy of predicted forest physiognomic types from support vector machine classification from three distinct groups, the significant optical bands, the significant LiDAR metrics, and the significant radar bands. Grey shade shows which bands were used.

Sentinel-2 band 2	[Grey shaded box]	
Sentinel-2 band 3		
Sentinel-2 band 4		
Sentinel-2 band 5		
Sentinel-2 band 8		
Sentinel-2 band 11		
Sentinel-2 band 12		
Mean of CHM		
97 <sup>th</sup> percentile of CHM	[Grey shaded box]	
Sentinel 1 VH/VV		
PALSAR HH		
Accuracy (%)	72.67	78.48      80.05



**Figure 1.** Map of Wellington region showing areal extent of indigenous forest in dark green. Indigenous forest occurs primarily in the Tararua, Rimutaka, and Aorangi mountain ranges. Map grid is the New Zealand Transverse Mercator.





**Figure 2.** Map of forest physiognomic types following support vector machine (SVM) classification of spectral and structural information from Sentinel-1 and 2, PALSAR (Phased Array L-band Synthetic Aperture Radar) and Light Detection and Ranging (LiDAR) in transect across the Tararua ranges. Map grid is the New Zealand Transverse Mercator. Broadleaved-podocarp (olive); Beech-broadleaved-podocarp (dark green); Beech-broadleaved (green-blue); Broadleaved (lime); Beech (blue).

Table 4 shows the areas of each forest physiognomic type as mapped in the Wellington region. The areas of the physiognomic types by height class (according to the 97th percentile of the CHM), short (< 20 m), medium (20–30 m), and tall (>30 m). Beech-broadleaved forests are the most common, with 63 thousand ha, and Broadleaved-podocarp forests are the second most common, with 35 thousand ha. Beech forests and Beech-broadleaved-podocarp forests are less common, both with areas of 16 thousand ha. Broadleaved forests are uncommon, with only 8 thousand ha. Podocarp forests do not exist in the Wellington region, (but do in other regions of New Zealand). Table 5 shows the common forest alliances within the physiognomic types in the Wellington region according to the forest plot data (note that no forest plots occur on the relatively uncommon broadleaved forest).

**Table 4.** Areas (ha) of physiognomic types of indigenous forest in the Wellington region by height class. Height classes are defined by the 97th percentile of CHM: short (<20 m); medium (20–30 m); and tall (>30 m).

Forest Physiognomic Type	Short	Medium	Tall	Total (ha)
Podocarp	0	0	0	0
Broadleaved-podocarp	6886	25,032	3985	35,902
Beech-broadleaved-podocarp	0	9929	6418	16,347
Beech-broadleaved	9815	52,593	862	63,270
Broadleaved	5733	2869	113	8715
Beech	8454	8237	106	16,797
Total indigenous forest				141,031
Total land area				811,727

**Table 5.** Common forest alliances in the Wellington region.

Physiognomic Types	Name of Forest Alliance	Number of Plots
Beech-broadleaved forest	Kāmahi-hardwood forest	7
	Silver beech-broadleaf forest	3
	Silver beech-red beech-kāmahi forest	16
Beech-broadleaved-podocarp forest	Kāmahi-Southern rata forest and tall shrubland	
	Pepperwood-hardwood forest and successional shrubland	12
	Kāmahi forest	27
	Kāmahi-silver fern forest	8
Beech forest	Black/mountain beech forest (subalpine)	
	Black/mountain beech – silver beech forest/subalpine shrubland	1
	Black/mountain beech forest	10
	Silver beech – red beech – black/mountain beech forest	1
	Silver beech forest with mountain lacebark and weeping matipo	1
	Hard beech – kāmahi forest	
Broadleaved-podocarp forest	Kāmahi-podocarp forest	1
	Mahoe forest	12
	Tawa forest	12
	Silver fern – mahoe forest	3
	Pepperwood – fuchsia – broadleaf forest	13
	Matai forest	
	Towai – tawa forest	

Height classes differ in their distribution among the physiognomic types. In Broadleaved-podocarp forest, medium is the most common height class, but short and tall are also common. In Beech-broadleaved-podocarp, medium and tall are common height classes, but there is no short. In Beech-broadleaved forest, medium is the most common height class, but short is also common, and there is very little tall. In Broadleaved forest, short and medium are common height classes, and there

is very little tall. Likewise, in Beech forest, short and medium are common height classes, and there is very little tall.

## 5. Discussion

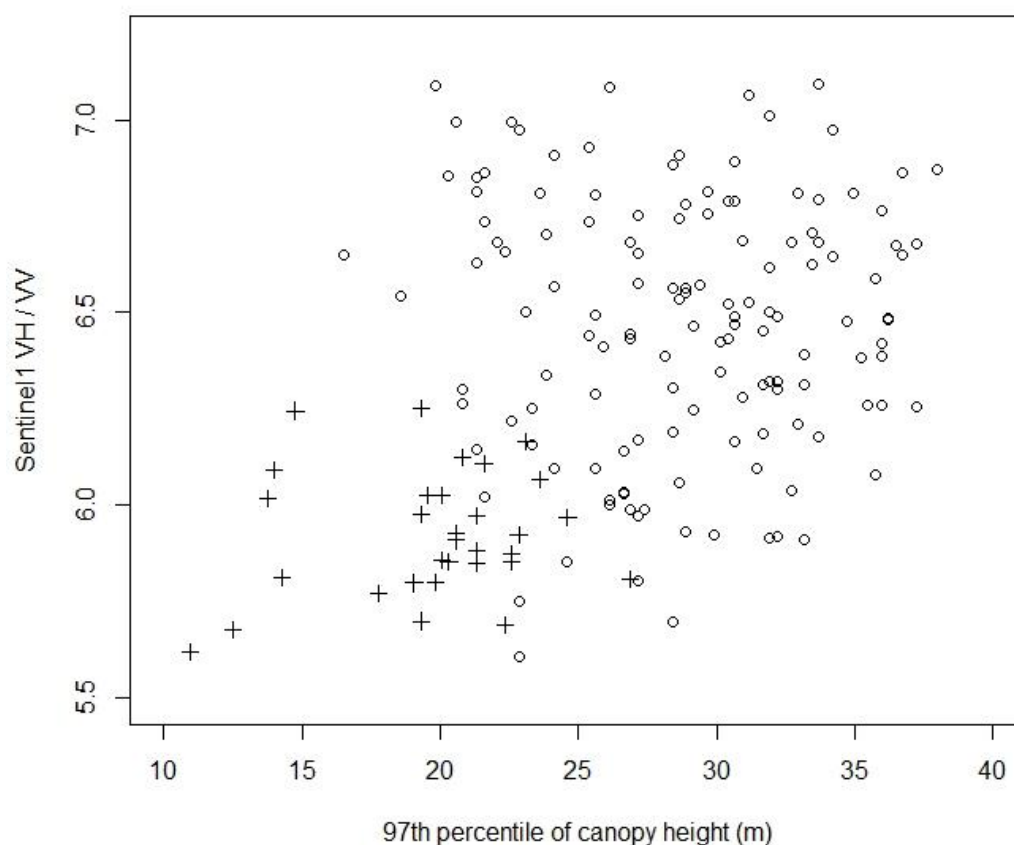
The 80.5% accuracy (rounded up from 80.48%) of the Wellington map of forest physiognomic types is a significant improvement over the 50.0% accuracy of a current national map of forest physiognomic types, EcoSat Forests [10] (published at a scale of 1:750,000), whose accuracy in the Wellington region we assessed using the NVS forest plots. EcoSat Forests were derived from a combination of Landsat imagery and the vegetation component of the New Zealand Land Resource Inventory (NZLRI). The Landsat imagery was used to bring the NZLRI vegetation up to date (i.e., to 2002/2003 from the 1970s) and to define forest boundaries within land units. The forest codes in the NZLRI do not have the mixed classes, Beech-broadleaved-podocarp forest and Beech-broadleaved forest, which are common in the Wellington region, except if they occupy separate parts of a land unit. It is not surprising, therefore, that there is a significant improvement in the method presented in this paper, which maps mixed forest physiognomic types directly. This was one of the reasons that EcoSat Forests was published at a scale of 1:750,000. However, it has been used at a range of scales, including 1:50,000, because it is the only national digital map of forest physiognomic types with reasonable currency (~ 2000).

While the accuracy of mapping forest physiognomic types is much improved with the method presented in this paper, it is still not high enough for some forest management applications, such as disease management. In Table 6, we list major management applications for indigenous forest mapping and assess (by canvassing expert opinion) whether the 80.5% classification accuracy would make the map of forest physiognomic types fit for purpose. Fitness of purpose is assigned a 'Yes' if a significant contribution is made to the application and 'No' if a minor contribution is made. Fitness for purpose is assigned 'Yes/no' if a significant contribution is made to some aspects of the application while a minor contribution is made to other aspects. The method presented in this paper (labelled 'Forest Physiognomic Types') is fit for purpose for four of the management applications, which compares favourably with one for EcoSat forests and none for LCDB. However, the method presented here has been applied to only one region, Wellington, and so EcoSat forests and LCDB will continue to be used for national applications until a national map of forest physiognomic types is available.

The Wellington region was chosen as the study area because it has a large proportion of indigenous forest and is one of the two regions in New Zealand with complete LiDAR coverage. Several other regional councils are currently acquiring regional LiDAR. The New Zealand government has recently recognised the utility of LiDAR data by announcing a nineteen-million-dollar funding boost for regional acquisition of LiDAR. All New Zealand may have LiDAR coverage within a few years, making a national map of forest physiognomic types possible using the method presented in this paper. There may be regional differences in the SVM classifier due to a gradual change in climate from north to south and also different soils and topography, but there is a good spread of vegetation plots throughout New Zealand that may be used to refine regional calibrations of the classifier.

The addition of structural information to optical data improves the classification of forest types. Indeed, even with just the ratio of Sentinel-1 VH over VV band and the 97% percentile of canopy height band, there is nearly complete discrimination between Broadleaved-podocarp and Beech forests (Figure 3). This is due to the very different canopy structure of these two physiognomic types. Broadleaved-podocarp forest commonly has tall podocarp trees in a sparse upper canopy (typically over 25 m tall), with a lower, 15-m tall secondary canopy. The canopy surface appears rough and heterogeneous and varies markedly in height. In contrast, Beech forest commonly has a smooth upper canopy with most trees reaching to the same upper canopy level at 20 m. Therefore, we would expect the ratio of volume scattering to surface scattering to be higher for the Beech forest than for the Broadleaved-podocarp forest, as seen in Figure 3.





**Figure 3.** Plot of Sentinel-1 ratio of VH/VV versus 97th percentile of canopy height for Beech forest (crosses) and Broadleaved-podocarp forest (circles).

**Table 6.** Fitness for purpose of method presented in this paper (labelled “Forest Physiognomic Types”) for major management applications for indigenous forest in New Zealand, compared with EcoSat Forests and Land Cover Database (LCDB).

Application	EcoSat Forests	LCDB	Physiognomic Types
Biodiversity management	Yes	Yes/No	Yes
Carbon inventory	No	No	Yes
Weed control	No	No	No
Pest control	Yes/No	No	Yes
Ungulate management	Yes/No	No	Yes
Disease management	No	No	Yes/No

Support Vector Machine classification has handled the addition of structural metrics to optical data satisfactorily and robustly, with the accuracy of training and test data being the same. It compares favourably to other classifiers which achieved lower classification accuracies in the R Caret package, such as Random Forests (71.5%), Extreme Random Trees (78.0%), Neural Networks (73.8%), Classification Trees (C5.0) (73.8%), Classification Trees (CART) (62.5%), Boosted Regression Trees (73.8%), Linear Discriminant Analysis (77.3%), K-Nearest Neighbour (70.5%), and Parallelepiped Classifier (53.3%). While Extreme Random Trees handled the addition of structural metrics well, as did the Support Vector Machine classifier, it was not robust – the accuracy of training data (100%) and test data (78.0%) was very different. Other classifiers exhibiting a lack of robustness were Random Forests and Classification Trees (both C5.0 and CART). The robustness of the Support Vector Machine is desirable when it is applied in regions where training data are sparse.

It would be possible to further increase the mapping accuracy of our method by adding environmental ancillary information, such as elevation from a DEM (Digital Elevation Model),

or a soil map [49]. However, we have chosen to limit the method to using direct measures of vegetation, rather than indirect environmental drivers. This is because accuracy is not everything—the ability to apply the method in other regions with minimum effort is also important, as is the ability to interpret results. When ancillary information is used, the complexity of the classifier increases, and overfitting becomes more likely when ground data are limited. Extension of the classifier to other areas without extensive ground data then becomes risky. Artefacts can also appear when forest boundaries are created by environmental thresholds, such as an elevation contour. While we prefer not to use environmental ancillary information, because we are aiming to produce a nationally applicable method, for applications in specific localities, it might be useful, even necessary. We will however, investigate the use of temporal spectral signatures, once the length of record of Sentinel-2 is sufficiently long—to further increase mapping accuracy and associated contribution to management applications.

## 6. Conclusions

Radar and canopy height metrics, which are related to vegetation structure, may be combined with optical imagery to improve the accuracy of forest classification. We combined C-band Sentinel-1 (VH/VV) radar, LiDAR derived mean canopy height, and 97th percentile of canopy height, with Sentinel-2 optical imagery to improve the mapping accuracy of an SVM classifier for mapping forest physiognomic types from 73.1% to 80.5%. The optical bands found useful for mapping these physiognomic types were blue ( $490 \pm 50$  nm), green ( $560 \pm 25$  nm), red ( $665 \pm 20$  nm), near infrared ( $835 \pm 65$ ), short-wave infrared ( $1610 \pm 70$  nm), and short-wave infrared ( $2185 \pm 120$  nm). The 80.5% classification accuracy of forest physiognomic types achieved by the combination is sufficiently high to contribute to several management applications for indigenous forest including biodiversity management, carbon inventory, pest control, ungulate management, and disease management.

**Author Contributions:** Conceptualization, J.D.; Data curation, S.W.; Formal analysis, J.D., J.Z., S.W. and M.S.; Methodology, J.D., J.Z. and J.S.; Resources, D.P.; Software, J.S.; Validation, J.D., J.Z. and J.S.; Writing—original draft, J.D., J.Z. and S.W.; Writing—review & editing, J.D. and D.P.

**Funding:** This research was funded by Ministry of Business, Innovation and Employment, contract number C09X1709.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wardle, P. *Vegetation of New Zealand*; Cambridge University Press: Cambridge, UK, 1991.
2. Dymond, J.R.; Ausseil, A.G.E.; Peltzer, D.A.; Herzig, A. Conditions and trends of ecosystem services in New Zealand—A synopsis. *Solut. J.* **2014**, *5*, 38–45.
3. Gordon, D.P. New Zealand's genetic diversity. In *Ecosystem Services in New Zealand—Conditions and Trends*; Dymond, J.R., Ed.; Manaaki Whenua Press: Lincoln, New Zealand, 2013.
4. Cockayne, L. Monograph on the New Zealand beech forests. Part 1. In *The Ecology of the Forests and Taxonomy of the Beeches*; Government Printer: Wellington, New Zealand, 1926.
5. Elliot, G.; Kemp, J. Large-scale pest control in New Zealand beech forests. *Ecol. Man. Restor.* **2016**, *17*, 200–209. [[CrossRef](#)]
6. New Zealand Forest Service. Forest Service Mapping Series 6. North Island and South Island. Available online: <https://koordinates.com/layers/?q=fsms6> (accessed on 15 August 2019).
7. Newsome, P.F.J. The vegetative cover of New Zealand. In *Water and Soil Miscellaneous Publication 112*; Water and Soil Directorate, Ministry of Works and Development: Wellington, New Zealand, 1987.
8. Wiser, S.; Thomson, F.; de Cáceres, M. Expanding an existing classification of New Zealand vegetation to include non-forested vegetation. *N. Z. J. Ecol.* **2016**, *40*, 160–178. [[CrossRef](#)]
9. Wiser, S.K.; de Cáceres, M. New Zealand's plot-based classification of vegetation. *Phytocoenologia* **2018**, *48*, 153–161. [[CrossRef](#)]



10. Shepherd, J.D.; Ausseil, A.-G.; Dymond, J.R. *EcoSat Forests: A 1,750,000 Scale Map of Indigenous Forest Classes in New Zealand*; Manaaki Whenua Press: Lincoln, New Zealand, 2005; Available online: <https://iris.scinfo.org.nz/> (accessed on 15 August 2019).
11. Dymond, J.R.; Shepherd, J.D.; Newsome, P.F.; Belliss, S. Estimating change in areas of indigenous vegetation cover in New Zealand from the New Zealand Cover Database (LCDB). *N. Z. J. Ecol.* **2017**, *41*, 56–64. [[CrossRef](#)]
12. Persson, M.; Lindberg, E.; Reese, H. Tree species classification with multi-temporal Sentinel-2 data. *Remote Sens.* **2018**, *10*, 1794. [[CrossRef](#)]
13. Zhu, L.; Suomalainen, J.; Liu, J.; Hyypä, J.; Kaartinen, H.; Haggren, H. A review: Remote sensing sensors. In *Multi-Purposeful Application of Geospatial Data*; Rustamov, R.B., Hasanova, S., Zeynalova, M.H., Eds.; Intechopen: London, UK, 2018; pp. 19–42.
14. Jain, M.; Srivastava, A.K.; Singh, B.; Joon, R.K.; McDonald, A.; Royal, K.; Lisaius, M.C.; Lobell, D.B. Mapping smallholder wheat yields and sowing dates using micro-satellite data. *Remote Sens.* **2016**, *8*, 860. [[CrossRef](#)]
15. Ghuffar, S. DEM generation from multi satellite PlanetScope imagery. *Remote Sens.* **2018**, *19*, 1462. [[CrossRef](#)]
16. Zhu, Z.; Woodcock, C.E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **2014**, *144*, 154–171. [[CrossRef](#)]
17. Gascon, F.; Bouzinac, C.; Thépaut, O.; Jung, M.; Francesconi, B.; Louis, J.; Lonjou, V.; Lafrance, B.; Massera, S.; Languille, F.; et al. Copernicus Sentinel-2A calibration and products validation status. *Remote Sens.* **2016**, *9*, 584. [[CrossRef](#)]
18. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Meygret, A.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [[CrossRef](#)]
19. Attema, E.; Snoei, P.; Davidson, M.; Floury, N.; Levrini, G.; Rommen, B.; Rosich, B. The European GMES Sentinel-1 Radar Mission. In Proceedings of the IGARSS 2008—2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 7–11 July 2008; pp. 1–97.
20. Clewley, D.; Whitcomb, J.; Moghaddam, M.; McDonald, K.; Chapman, B.; Bunting, P. Evaluation of ALOS PALSAR Data for High-Resolution Mapping of Vegetated Wetlands in Alaska. *Remote Sens.* **2015**, *7*, 7272–7297. [[CrossRef](#)]
21. Lefsky, M.A.; Cohen, W.B.; Acker, S.A.; Parker, G.G.; Spies, T.A.; Harding, D. Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Remote Sens. Environ.* **1999**, *70*, 339–361. [[CrossRef](#)]
22. Regional High-Resolution DEM Now on LDS, Wellington Regional Government GIS Group. Available online: <http://mapping.gw.govt.nz/News06.htm> (accessed on 15 August 2019).
23. Sothe, C.; de Almeida, C.M.; Liesenberg, V.; Schimalski, M.B. Evaluating Sentinel-2 and Landsat-8 data to map successional forest stages in a subtropical forest in Southern Brazil. *Remote Sens.* **2017**, *9*, 838. [[CrossRef](#)]
24. Deng, S.; Katoh, M.; Xiaowei, Y.; Hyypä, J.; Gao, T. Comparison of Tree Species Classifications at the Individual Tree Level by Combining ALS Data and RGB Images Using Different Algorithms. *Remote Sens.* **2016**, *8*, 1034. [[CrossRef](#)]
25. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
26. Adam, E.; Mutanga, O.; Odindi, J.; Abdel-Rahman, E.M. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: Evaluating the performance of random forest and support vector machines classifiers. *Int. J. Remote Sens.* **2014**, *35*, 3440–3458. [[CrossRef](#)]
27. Fassnacht, F.E.; Latifi, H.; Sterenczak, K.; Modzelewska, A.; Lefsky, M.; Waser, L.T.; Straub, C.; Ghosh, A. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **2016**, *186*, 64–87. [[CrossRef](#)]
28. Liu, Y.; Gong, W.; Hu, X.; Gong, J. Forest type identification with random forest using sentinel-1A, Sentinel 2-A, multi-temporal Landsat-8 and DEM data. *Remote Sens.* **2018**, *10*, 946. [[CrossRef](#)]
29. Piazza, A.G.; Vibrans, A.C.; Liesenberg, V.; Refosco, J.C. Object-oriented and pixel-based classification approaches to classify tropical successional stages using airborne high-spatial resolution images. *GISci. Remote Sens.* **2016**, *53*, 206–226. [[CrossRef](#)]
30. Féret, J.; Asner, G.P. Tree species discrimination in tropical forests using airborne imaging spectroscopy. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 73–84. [[CrossRef](#)]

31. Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. A review of supervised object-based land-cover image classification. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 277–293. [CrossRef]
32. Clark, M.L.; Roberts, D.A.; Clark, D.B. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sens. Environ.* **2005**, *96*, 375–398. [CrossRef]
33. Clark, M.L.; Roberts, D.A. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sens.* **2012**, *4*, 1820–1855. [CrossRef]
34. Mickelson, J.G.; Civco, D.L.; Silander, J.A. Delineating forest canopy species in the northeastern United States using multi-temporal TM imagery. *Am. Soc. Photogramm. Remote Sens.* **1998**, *64*, 891–904.
35. Wolter, P.; Mladenoff, D.J. Improved Forest Classification in the Northern Lake States Using Multi-Temporal Landsat Imagery. *Photogramm. Eng. Remote Sens.* **1995**, *61*, 1129–1143.
36. Zhu, X.L.; Liu, D.S. Accurate mapping of forest types using dense seasonal Landsat time-series. *ISPRS J. Photogramm. Remote Sens.* **2014**, *96*, 1–11. [CrossRef]
37. Hill, R.A.; Wilson, A.K.; George, M.; Hinsley, S.A. Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Appl. Veg. Sci.* **2010**, *13*, 86–99. [CrossRef]
38. Dymond, J.R.; Shepherd, J.D. The spatial distribution of indigenous forest and its composition in the Wellington region, New Zealand, from ETM+ satellite imagery. *Remote Sens. Environ.* **2004**, *90*, 116–125. [CrossRef]
39. Definiens. *eCognition Version 5 Object Oriented Image Analysis User Guide*; Definiens: Munich, Germany, 2005.
40. Bunting, P.; Clewley, D.; Lucas, R.M.; Gillingham, S. The remote sensing and GIS software library (RSGISLib). *Comput. Geosci.* **2014**, *62*, 216–226. [CrossRef]
41. New Zealand eScience Infrastructure. Available online: <https://www.nesi.org.nz/services/high-performance-computing> (accessed on 15 August 2019).
42. Shuttle Radar Topography Mission. Available online: <https://www2.jpl.nasa.gov/srtm/> (accessed on 15 August 2019).
43. Müller, M.U.; Shepherd, J.D.; Dymond, J.R. Support vector machine classification of woody patches in New Zealand from synthetic aperture radar and optical data, with LiDAR training. *J. Appl. Remote Sens.* **2015**, *9*, 095984. [CrossRef]
44. Bunting, P.; Armston, J.; Clewley, D.; Lucas, R. The sorted pulse data software library (SPDLib): Open source tools for processing LiDAR data. In Proceedings of the SilviLaser 2011, 11th International Conference on LiDAR Applications for Assessing Forest Ecosystems, University of Tasmania, Hobart, Australia, 16–20 October 2011; pp. 1–11.
45. Zörner, J.; Dymond, J.R.; Shepherd, J.D.; Wiser, S.K.; Jolly, B. LiDAR-based regional inventory of tall trees—Wellington, New Zealand. *Forests* **2018**, *9*, 702. [CrossRef]
46. Wiser, S.K.; Bellingham, P.J.; Burrows, L. Managing biodiversity information: Development of the National Vegetation Survey Databank. *N. Z. J. Ecol.* **2001**, *25*, 1–17.
47. Hurst, J.M.; Allen, R.B. *A Permanent Plot Method for Monitoring Indigenous Forests—Field Protocols*; Landcare Research: Lincoln, New Zealand, 2007; Volume 3, pp. 154–196.
48. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. Available online: <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (accessed on 15 August 2019).
49. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [CrossRef]

