National forest owner survey and resource inventory of alternative species

Stage 3: Mapping alternative species in East Coast

Authors: Vega Xu and Bruce Manley

Research Providers: University of Canterbury

Publication No: SWP-T159

Date: Dec 2022

Leadership in forest and environmental management, innovation and research

Executive Summary

This study used a remote sensing approach to classify alternative species of small-scale plantations in the East Coast wood supply region (i.e. Gisborne District). With additional truthing data, the approach achieved 92.9% overall classification accuracy. Douglas-fir and *Eucalyptus* were the two most accurately classified alternative species categories, with producer's accuracies of 97.2 and 94.0% respectively. The most important input variable selected for the classification was DEM (Digital Elevation Model), suggesting that elevation plays an important role in differentiating plantation species.

When applying the classification to the East Coast region, overall 4,582 ha of small-scale alternative species were mapped and the most common alternative species categories are Douglas-fir and *Eucalyptus*, accounting for 35% and 30% of the total small-scale alternative species resources. Acacia and poplar are the least common alternative species identified, with 72 ha and 59 ha estimated respectively. When aggregated with the area provided by the large-scale owners, in total 5,353 ha of alternative species were estimated in the East Coast region. This is 780 ha (17%) more than the NEFD-reported area. The area of cypress, other softwoods and hardwoods are similar to the NEFD area. However, Douglas-fir was 245 ha (12%) less than the NEFD area and the estimated *Eucalyptus* area was three times more than the NEFD area. Overall, it appears that NEFD underestimates the total area of small-scale alternative species.

TABLE OF CONTENTS

Project Introduction

The Stage 1 objective of this project was to identify a suitable methodology for a NZ-wide survey and inventory of alternative species forests and their owners by undertaking a pilot study in Hawke's Bay Region. This objective was achieved, by mapping the alternative species resource down to 0.1 ha in the Hawke's Bay Region, and identifying a significant proportion of the forest owners via the application of the LINZ cadastral layer. The work confirmed that there are significant differences between National Exotic Forest Description (NEFD) data for alternative species in Hawke's Bay and the actual forest resource on the ground. Whereas the NEFD estimated 3190 ha of alternative species, the study found 3794 ha of area above the 1 ha minimum applied in the NEFD.

Stage 2a objectives were to validate and extend the small-scale owners' data gathered in Stage 1 and generate information about the Hawke's Bay alternative species resource. A survey on small-scale alternative species was delivered to small-scale owners identified from Stage 1, however, the response rate was very poor and the survey did not provide much information on the alternative species resource.

The objective of Stage 2b was to develop an automated mapping approach to classify alternative species based on input data received from corporate owners. The mapping approach achieved an overall classification accuracy of 92.8%. Douglas-fir and *Eucalyptus* were the two most accurately classified alternative species categories, with producer's accuracy of over 90%. The most important input variable selected for the classification was DEM (Digital Elevation Model), suggesting that elevation plays an important role in differentiating plantation species. It was found that the accuracy of species classification highly depends on the availability of truthing data.

Stage 3 objective: the classification method developed in stage 2b on the Hawke's Bay resources will be applied to the East Coast region. Additional truthing data with known locations and species will be collected from PSPs and NZDFI trials, and then be used in the classification of alternative species in the East Coast region.

Stage 3: Introduction

The key objective of stage 3 was to develop a map of alternative species in the East Coast wood supply region (i.e. Gisborne District). Random Forest (RF) classifier will continue to be applied using 10 m resolution Sentinel imagery due to its robust performance proved in stage 2b. Truthing data with known locations and species are required to perform species classification, and the data should be representative and cover a wide range of age classes and site conditions. Large-scale forest owners generally have this data in digital format for management purposes, whereas smaller-scale owners rarely have such good digital records. In stage 2b, the large-scale owners for East Coast provided spatial data for 771 ha of alternative species, mainly consisting of eucalyptus, Douglas-fir and pine (non-radiata) species (Table 1). This data was used as truthing data for stage 3. However, in stage 2b, it was found that the classification results highly depended on the amount of truthing data. Therefore, we expanded the truthing data by incorporating PSP data for alternative species for the stage 3 mapping.

Table 1: Area of alternative species provided by large-scale owners in East Coast.

Species	Area (ha)
Acacia	26
Cypress	90
Douglas-fir	138
Eucalyptus	196
Larch	30
Native	3
Other exotics	63
Other mixed	5
Pine	153
Redwood	67
Total	**771**

The 2021 National Exotic Forest Description (NEFD) reported 4,573 ha of alternative species in the East Coast wood supply region. According to our experience with this project, NEFD may not be accurate and the numbers do not provide the spatial location of the alternative species.

Table 2: Area of alternative species in East Coast wood supply region as reported in 2021 NEFD.

Alternative species	NEFD Area (ha)
Douglas-fir	1,987
Cypress species	282
Other softwoods	887
Eucalypt	492
Other hardwoods	925
Total	**4,573**

This project aims to map the alternative species within the East Coast wood supply region with RF classifier and additional truthing data. Specifically, the project involved three components: pre-defining the alternative species boundaries manually with aerial photos, collecting additional truthing data and preform species classification. The mapped area was consequently be compared with the NEFD area.

Methods

Pre-defining forest boundary

Pre-defining the geographic boundaries of alternative species is required to define the extent of classification. Without the pre-defined boundaries, the classification approach tends to map other land covers as alternative species plantations due to a similar spectral signature. An operator was trained to manually delineate the boundary of alternative species that are sized over 0.5 ha in East Coast using 0.3 m aerial photos downloaded from LINZ.

Collecting additional truthing data

As discovered from stage 2b, the truthing data for East Coast only covers 771 ha of alternative species, the truthing data in Central North Island (CNI) and Hawke’s Bay were also used as additional truthing data.

In addition, PSP data including the coordinates of the plot, plot size, species and establishment year were acquired from Scion. In total 847 plots were received including Douglas-fir, cypress, *Eucalyptus*, acacia and redwood, which covers the whole country. There were 26 plots for East Coast (redwood, eucalyptus and cypress). NZDFI also provided the GIS boundaries of their trials which include detailed *Eucalyptus* species and establishment year information.

Within the provided GIS boundaries provided, circular plots with a maximum of 50 m radius were automatically and randomly generated which were then used as the sample data for species classification in order to reduce classification time. The data was then randomly split into 70% training and 30% validation dataset. A summary of the size of truthing data for each target classification class is described in Table 3.

Table 3: Description of training and validation data for each species class. Each pixel represents a 10x10 m grid. Other species include other alternative species that are not listed in the table such as cedar and willow. Radiata pine samples were manually added as place holders in the classification. Other pines are pine species other than radiata pine.

Species	Training	Validation	Total
Acacia	2,281	977	3,258
Cypress	10,384	4,449	14,833
Douglas-fir	41,850	17,935	59,785
Eucalyptus	17,049	7,306	24,355
Larch	2,229	954	3,183
Other pine	6,135	2,628	8,763
Other species	7,902	3,386	11,288
Poplar	1,592	681	2,273
Radiata	27,622	11,837	39,459
Redwood	5,606	2,402	8,008
Total	**122,650**	**52,555**	**175,205**

Remote sensing data

The national Sentinel-2 mosaic was processed by Manaaki Whenua - Landcare Research based on workflow developed by Shepherd, et al. (2020) and distributed by the Ministry for the Environment (MfE), New Zealand. The image product is a 10 m, a ten-band multispectral, cloud-minimised mosaic of multiple Sentinel-2A and -2B satellite images over New Zealand and was acquired from late 2021 to early 2022 (Table 4). The mosaic went through pan-sharpening, atmospheric and bidirectional reflectance distribution function correction, cloud clearing and minimising process.

Table 4: Bands included in Sentinel-2 image mosaic

Band	Band Name	Short Name	Wavelength (nm)
2	Blue	B	490
3	Green	G	560
4	Red	R	665
5	Red Edge 1	RE705	705
6	Red Edge 2	RE740	740
7	Red Edge 3	RE783	783
8	Near Infrared wide	NIR842	842
8A	Near Infrared narrow	NIR865	865
11	Short Wave Infrared 1	SWIR1610	1610

12	Short Wave Infrared 2	SWIR2190	2190

Vegetation indices (VIs), which are the spectral transformation of two or more spectral bands, are useful in detecting variations in spectral response to foliage colour and have considerable advantages in cellular structure evaluation, stress prediction, moisture content estimate, pigment content detection, and stress estimation (Immitzer et al., 2019). In total, 33 vegetation indices, which are sensitive to vegetation properties and have been previously used in vegetation classification studies (Grabska et al., 2019; Immitzer et al., 2019; Ye et al., 2021), were extracted from the Sentinel-2 mosaic (Appendix 1).

Textural features are mainly related to the variability of stand density, forest type (broadleaved, coniferous), crown size, crown closure, crown form, and crown closure (Fassnacht et al., 2016). They can considerably enhance the classification accuracy when combined with spectral features (Mallinis et al., 2008). For this study, after performing Principle Component Analysis (PCA) on the Sentinel-2 mosaic, a 3 by 3 window size was used to calculate the values of the Grey Levels Co-Occurrence Matrix (GLCM) (mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation). To minimise computation and data, the first principle band was chosen for the GLCM; the majority of the data variance could be kept by performing PCA on the original image.

Phenological features were derived from analysing the temporal variation of Enhanced Vegetation Index-2 (EVI2) using Sentinel-2 data collected from 1 January 2019 to 31 December 2020 in Google Earth Engine (GEE). EVI2 was chosen because it is one of the most commonly used VIs for phenological studies, as reviewed by Caparros-Santiago et al. (2021). It was developed by Jiang et al. (2008) to address the saturation issue of the Normalized Difference Vegetation Index (NDVI) in areas with high biomass and to avoid using the blue band, which lacks vegetation characteristic information in the calculation (Wang et al., 2018). Three seasonal metrics, amplitude (AMP), phase (PH) and mean EVI2 of the period, were extracted. Phase measures the length of the change's time window, whereas amplitude shows the size of the shift relative to a baseline.

In addition, a Digital Elevation Model (DEM) was retrieved from Land Information New Zealand (Land Information New Zealand (LINZ), 2020) and was resampled to 10 m to be consistent with the rest of the input features. In total, 55 features were extracted using remote sensing software ENVI version 5.6 (ENVI, 2021).

Species classification

Random forest is a machine learning algorithm applied widely in image classification because of its high prediction accuracy and the ability to handle high-dimensional data. The classifier is an ensemble of independent individual decision trees, each individual decision tree in the classifier casts a vote for the class that should be applied to the given sample, and the class that receives the most votes wins (Breiman, 2001). The algorithm does not require distributional assumption and is less sensitive to the number of input variables and overfitting (Fassnacht et al., 2016). Pelletier et al. (2016) compared classification algorithms and concluded random forest is most robust in mapping land cover over large areas by producing the highest classification accuracy with the shortest training time, as well as being less affected by parametrisation and the number of training samples. Therefore, pixel-based classification with the random forest classifier was applied using the "randomForest" package (Liaw & Wiener, 2002) in statistical package R (R Core Team, 2013).

Due to high dimensional input features and target species classes, a feature selection process using the "VSURF" package (Genuer et al., 2015) was applied to eliminate redundant variables and reduce computation time for classification. Based on findings from Speiser et al. (2019), VSURF outperformed other feature selection methods for random forest classification. After the classification, a majority filter (with 8 x 8 neighbours) was applied to the classification image to minimise the occurrence of small isolated pixels.

The accuracy of classification was assessed by using the most common approach - the confusion matrix (Congalton, 2001), which compares the classified and truth species classes based on the validation dataset. Measures such as the overall accuracy, the producer's and user's accuracies for individual classes were calculated. The overall accuracy indicates the proportion of pixels that were correctly classified out of all the truth pixels. The producer's accuracy, which is related to omission error, reflects the probability of a species class being correctly classified. The user's accuracy relates to the commission error, which represents the probability that a pixel classified into a given species actually represents that species on the ground.

Area comparison

The output classified image was clipped to the extent of the small-scale alternative species in the East Coast region which was mapped manually so that the area of each species class can be calculated within the mapped extent. The areas were then aggregated with the summary of the large-scale alternative species to provide a full area description of the alternative species in East Coast.

Results and Discussion

Spectral signature of species

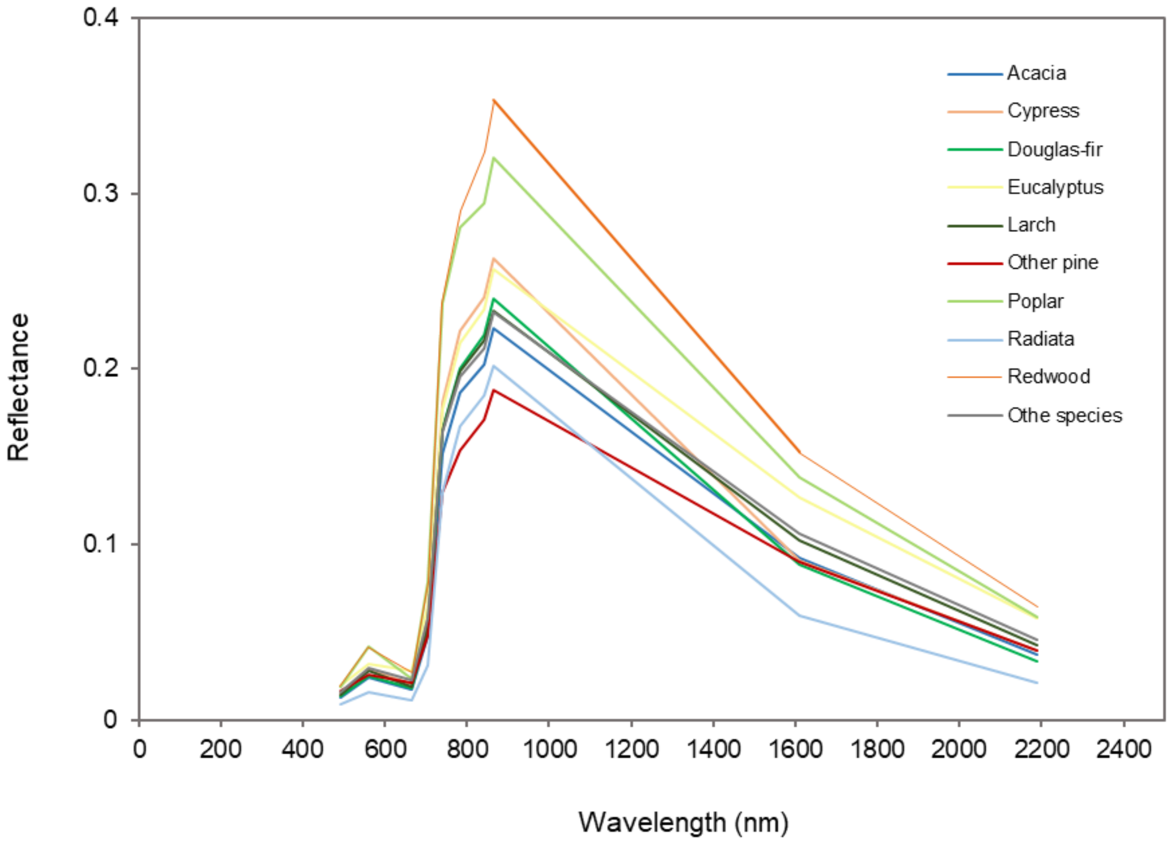

Figure 1: Spectral signature of different species on Sentinel-2 imagery

Different land covers absorb, emit and reflect different wavelengths of the electromagnetic spectrum. A predictive model known as "spectral signatures" was created using multivariate statistical algorithms using truthing data and multi-spectral satellite data for the same sites in order to categorise the satellite image into different types of land cover (Laborte et al., 2010).

Prior to species classification, the spectral signature of each tree species indicating how species' reflectance differs between the wavelength bands, was examined to understand the potential separability of different species (**Error! Reference source not found.**). The spectral signature suggests that generally, all tree species reflect similarly within the visible wavelength (400-700 nm) but illustrate the higher separation between the reflectance in the red edge and NIR spectrum (700-1300 nm). The SWIR spectrum (1300-2500 nm) also indicated some level of separation of reflectance. The spectral signature of all species showed a preliminary possibility of separating tree species at the RE, NIR and SWIR spectra. The reflectance characteristics of individual leaf components play the main role in how radiation interacts with vegetation. Chlorophyll, carotenoids, and anthocyanins, which are pigments found in leaves, absorb incident light to produce the majority of the visible spectrum's signal. Water absorption is the main factor in the NIR spectrum. Water has a major role in determining the reflectance in the SWIR region, although nitrogen and different types of carbon also contribute significantly to the reflectance (Asner, 1998).

Input Features

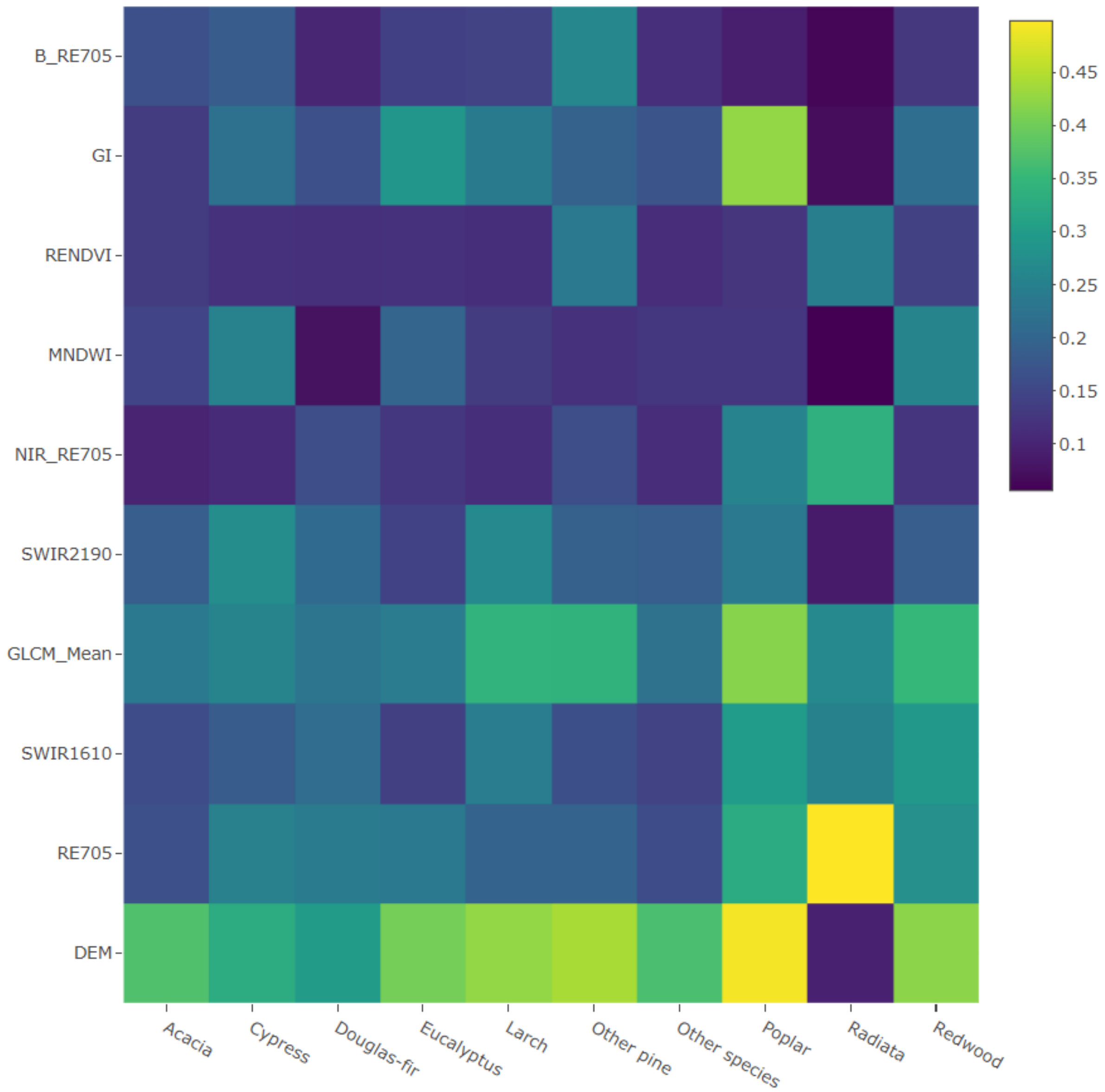

Figure 2: The importance score of the selected variables for each species class

After running VSURF variable selection process, ten out of 55 input variables were selected for species classification and they also contributed differently to the species classification (**Error! Reference source not found.**). According to the importance score of all variables, DEM was the most useful variable for classifying all minor species, suggesting that elevation plays an important role in differentiating plantation species. DEM was also found as the most important contributor to land cover and forest species classification in other studies (Ye et al., 2021; Zhang & Yang, 2020). Following DEM, one textural feature GLCM_Mean was also identified as an important variable. Ghosh and Joshi (2014) and Ye et al. (2021) also discovered that GLCM-Mean contributed to mapping forest mapping due to the its ability to capture texture data while the vegetation composition was complicated.

Vegetation Indices (VIs) combine the surface reflectance at two or more wavelengths to emphasise a specific characteristic of vegetation, such as photosynthetic activity and canopy structure. They enhance the sensitivity of spectral properties of vegetation while reducing spectral disturbance (Glenn et al., 2008). VIs describe the biochemical and physiological properties of vegetation that could contribute to the vegetation classification. Five out of the ten variables were vegetation indices (B_RE705, GI, RENDVI, MNDWI and NIR_RE705).

Two original bands, RE705 and SWIR1610 were also identified as useful variables, which shows the consistency of findings from the spectral signature of species. RE and SWIR bands were also identified as high-value bands for forest species mapping (Immitzer et al., 2016) and land cover classification (Schuster et al., 2012).

Classification results

The species classification with all 55 input variables achieved an overall accuracy of 0.929 (Table 5), indicating 92.9% of the validation pixels were correctly classified. Douglas-fir and eucalyptus were the two most accurately classified alternative species, with 97.2% and 94% of producer's accuracies respectively, and 92.4% and 91.1% of user's accuracies respectively. These two classes also contain more truthing data than other classes. On the other hand, acacia has the lowest producer's accuracy (62.1%), which is likely due to less truthing data.

All classes achieved high user's accuracies (over 90%) except other pine species, indicating 90% of the pixels classified actually represent these species in the real world.

The overall classification accuracy using the selected variables was the same as using all input variables (Table 6). The differences in the user's and producer's accuracies were also minimal. This indicates the redundancy of input variables when using all 55 variables. Therefore, the classification algorithm with selected variables was chosen to be applied to the whole study area, due to similar accuracy and reduced computation time.

The overall classification accuracy for classifying multiple tree species was comparable with other studies using Sentinel-2 imagery, e.g. Bolyn et al. (2018) classified 10 tree species with an overall accuracy of 88.9%, Persson et al. (2018) produced an overall accuracy of 88.2% for classifying five tree species in a Swedish forest, Grabska et al. (2019) achieved up to 92.38% overall accuracy for classifying four tree species.

Overall, this study successfully classified forest species in highly fragmented forests over a large geographic area. However, it is challenging to achieve high accuracies for certain tree species (e.g. acacia and other species). Similarly, Immitzer et al. (2016) also observed lower classification accuracies for those tree species which are either uncommon in the study area or within mixed stands.

Table 5: Confusion matrix of classification with all input features. It was produced based on 30% validation dataset. PA stands for producer's accuracy and UA stands for user's accuracy. Overall accuracy is 0.929 and kappa coefficient is 0.910.

	Reference											
Prediction	Acacia	Cypress	Douglas-fir	Eucalyptus	Larch	Other species	Other pine	Poplar	Radiata	Redwood	Total	UA
Acacia	607	24	13	14	1	1	10	0	0	1	671	0.905
Cypress	37	3925	40	32	38	26	66	3	22	75	4264	0.920
Douglas-fir	174	229	17441	147	57	130	436	20	96	145	18875	0.924
Eucalyptus	99	81	106	6871	20	74	159	18	56	55	7539	0.911
Larch	2	18	5	5	786	12	1	4	0	2	835	0.941
Other species	19	19	25	45	4	2334	43	5	10	13	2517	0.927
Other pine	19	68	166	51	31	36	2578	33	17	41	3040	0.848
Poplar	0	0	0	1	5	0	16	588	0	6	616	0.955
Radiata	15	52	87	107	4	13	61	0	11633	26	11998	0.970
Redwood	5	33	52	33	8	2	16	10	3	2038	2200	0.926
Total	977	4449	17935	7306	954	2628	3386	681	11837	2402	52555	
PA	0.621	0.882	0.972	0.940	0.824	0.888	0.761	0.863	0.983	0.848		0.929 (0.910)

Table 6: Confusion matrix of classification with 12 selected variables. It was produced based on 30% validation dataset. PA stands for producer's accuracy and UA stands for user's accuracy. Overall accuracy is 0.929 and kappa coefficient is 0.910.

	Reference											
Prediction	Acacia	Cypress	Douglas-fir	Eucalyptus	Larch	Other species	Other pine	Poplar	Radiata	Redwood	Total	UA
Acacia	637	18	13	16	2	2	5	0	2	1	696	0.915
Cypress	29	3975	60	33	42	28	66	3	22	65	4323	0.920
Douglas-fir	157	180	17378	146	62	129	386	19	83	159	18699	0.929
Eucalyptus	85	75	116	6827	23	88	140	20	50	56	7480	0.913
Larch	3	19	7	14	780	7	12	1	0	5	848	0.920
Other species	27	22	36	49	4	2319	52	4	3	12	2528	0.917
Other pine	23	68	163	50	20	38	2612	21	21	36	3052	0.856
Poplar	0	1	1	1	8	0	22	601	0	6	640	0.939
Radiata	12	51	103	132	3	13	64	1	11652	26	12057	0.966
Redwood	4	40	58	38	10	4	27	11	4	2036	2232	0.912
Total	977	4449	17935	7306	954	2628	3386	681	11837	2402	52555	
PA	0.652	0.893	0.969	0.934	0.818	0.882	0.771	0.883	0.984	0.848		0.929 (0.910)

Area Comparison

Table 7: Area summary of alternative species in East Coast wood supply region. The small-scale areas were estimated using the classification, and the large-scale areas were from survey.

Species	Small-scale (ha)	Large-scale (ha)	Total (ha)
Acacia	72	26	98
Cypress	191	90	281
Douglas-fir	1,604	138	1,742
Eucalyptus	1,388	196	1,584
Larch	196	30	226
Other pine	139	153	292
Other species	622	71	693
Poplar	59	0	59
Redwood	311	67	378
Total (ha)	**4,582**	**771**	**5,353**

The manual mapping resulted in 4,587 ha of over 0.5 ha alternative species in East Coast. After clipping the classification result to the boundary, 4,582 ha of alternative species were left suggesting 5 ha of forests were lost during the clipping process. Overall the classification suggests the most common alternative species are Douglas-fir and *Eucalyptus*, accounting for 35% and 30% of the total small-scale alternative species resources. Acacia and poplar are the least common alternative species identified, with 72 ha and 59 ha estimated respectively (Table 7).

Table 8: The total area of alternative species compared in East Coast wood supply region with NEFD 2021 area. The area estimated in this study include both large-scale and small-scale. Other species are aggregated due to different species class definition in NEFD.

Alternative species	NEFD Area (ha)	Area from this study over 0.5 ha (ha)	Area from this study over 1 ha (ha)
Douglas-fir	1,987	1,742	1,443
Cypress species	282	281	273
Eucalypt	492	1,584	1,325
Other softwoods	887	897	844
Other hardwoods	925	850	752
Total	**4,573**	**5,353**	**4,637**

When aggregated with the area provided by the large-scale owners, the total area of each species group can be obtained (Table 7). In total 5,353 ha of alternative species were estimated in the East Coast region. This is 780 ha (17%) more than the NEFD-reported area (Table 8). The area of cypress, other softwoods and hardwoods are similar to the NEFD area. However, Douglas-fir was 245 ha (12%) less than the NEFD area and the estimated eucalyptus was three times more than the NEFD area. Overall, it appears that NEFD underestimated the area of small-scale alternative species. However, the minimum forest size in the NEFD is 1 hectare whereas in this study a minimum size of 0.5 ha was adopted. If only over 1-hectare forests are included in the estimation, in total 4637 ha of alternative species are mapped in the East Coast region. That is 64 ha more than the NEFD-reported area.

Limitation and Future Research

This study applied a random forest classifier to automatically classify species within pre-defined alternative species boundaries. The region of interest, East Coast, did not contain enough truthing data so the truthing data collected from Stage2 (CNI and Hawke's Bay) were used.
In addition, PSP data and NZDFI data were also added to increase the size of the truthing data.

Similar to findings in Stage 2b, the larger the truthing dataset, the higher the classification accuracy. Douglas-fir and *Eucalyptus* contained more truthing data and they are the most accurately classified species. Species like acacia, other pine species and other alternative species, which are less representative and have less truthing data, were classified less accurately. However, it is challenging to acquire more truthing data for those species as there is not much area available.

Another limitation of the study is that pre-defining the geographic boundaries of alternative species is required to define the extent of classification. Without the pre-defined boundaries, the classification approach tends to map other land covers as alternative species plantations due to a similar spectral signature.

This approach using random forest classifier and Sentinel imagery performed well in the East Coast region, which again confirms the robustness of the approach. However, given the 10-m resolution, it is challenging to classify at the individual species level. There is an opportunity to perform classification using higher resolution imagery, such as Worldview imagery or UAV images to classify more detailed species. With truthing data for individual species and higher resolution imagery, it might be possible to classify individual species within the same genus (e.g. eucalyptus species). Being able to differentiate at individual species level would improve the usefulness of undertaking a national inventory.

Given the ongoing acquisition of LiDAR data under the LINZ National Elevation Programme, there is also an opportunity to improve the classification with LiDAR-derived data such as elevation and canopy height model. As LiDAR data is only available in the East Coast region but not for the CNI and Hawke's Bay region, this project was unable to utilise LiDAR in the classification. However, once the national coverage of LiDAR becomes available, LiDAR-derived metrics can be used and potentially improve the species classification. It also adds value by estimating stand variables (such as height and volume) of alternative species given that inventory data would be available for ground truthing.

Conclusions

This study confirms the usefulness of random forest classification with Sentinel imagery in classifying alternative species at a regional level and achieved high classification accuracies for most species. The results are similar to the findings in Stage 2. The classification accuracy of using a machine learning classifier highly depends on the availability of truthing data. In total, 4582 ha of small-scale alternative species were classified for East Coast and a majority of them are Douglas-fir and *Eucalyptus*. A truthing database can be built up and this approach could be applied to all regions of New Zealand.

Appendix

Appendix 1: Input variables used in Random forest classification

Abbreviation	Name
Spectral bands	
Blue	Blue band
Green	Green band
Red	Red band
RE705	Red Edge 705 nm
RE740	Red Edge 740 nm
RE783	Red Edge 783 nm
NIR842	Near Infrared 842 nm
NIR865	Near Infrared 865 nm
SWIR1610	Short-wave infrared 1610 nm
SWIR2190	Short-wave infrared 2190 nm
Textural	
GLCM_Mean	Local mean of Gray-Level Co-Occurrence Matrix (GLCM)
GLCM_Variance	Local variance of GLCM
GLCM_Homogeneity	GLCM Homogeneity
GLCM_Contrast	GLCM Contrast
GLCM_Dissimilarity	GLCM Dissimilarity
GLCM_Entropy	GLCM Entropy
GLCM_2ndMoment	GLCM 2nd Moment
GLCM_Correlation	GLCM Correlation
Phenology	
Mean EVI2	The average Enhanced Vegetation Index 2 (EVI2)
EVI2 phase	The phase of EVI2
EVI2 amplitude	The amplitude of EVI2
Topography	
DEM	Resampled 10 m Digital Elevation Model
Vegetation Indices	
EVI2	Enhanced Vegetation Index2
GEMI	Global Environmental Monitoring Index
GARI	Green Atmospherically Resistant Index
GCI	Green Chlorophyll Index
GI	Greenness Index
GNDVI	Green Normalised Difference Vegetation Index
LAI	Leaf Area Index
MCARI_I	Modified Chlorophyll Absorption Ratio Index – Improved
MNLI	Modified Non-Linear Index
MNDWI	Modified Normalised Difference Water Index
MSR	Modified Simple Ratio
MSAVI2	Modified Soil Adjusted Vegetation Index 2
MTVI_I	Modified Triangular Vegetation Index – Improved
NDVI	Normalised Difference Vegetation Index
OSAVI	Optimized Soil Adjusted Vegetation Index
RENDVI	Red Edge Normalised Difference Vegetation Index
REPI	Red Edge Position Index
RGRI	Red Green Ratio Index
RDVI	Renormalised Difference Vegetation Index
SAVI	Soil Adjusted Vegetation Index
NIR_R	Simple Ratio NIR/red
B_RE705	Simple Ratio blue/RE705
B_RE740	Simple Ratio blue/RE740
B_RE783	Simple Ratio blue/RE783
NIR_B	Simple Ratio NIR/blue
NIR_G	Simple Ratio NIR/green
N_RE705	Simple Ratio NIR/RE705
N_RE740	Simple Ratio NIR/RE740
N_RE783	Simple Ratio NIR/RE783
TCARI	Transformed Chlorophyll Absorption Reflectance Index
TVI	Triangular Vegetation Index

Abbreviation	Name
VARI	Visible Atmospherically Resistant Index
WDRVI	Wide Dynamic Range Vegetation Index

References

Asner, G. P. (1998, 1998/06/01/). Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sensing of Environment, 64*(3), 234-253. https://doi.org/10.1016/S0034-4257(98)00014-5

Bolyn, C., Michez, A., Gaucher, P., Lejeune, P., & Bonnet, S. (2018). Forest mapping and species composition using supervised per pixel classification of sentinel-2 imagery [Article]. *Biotechnology, Agronomy and Society and Environment, 22*(3), 172-187. https://doi.org/10.25518/1780-4507.16524

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Caparros-Santiago, J. A., Rodriguez-Galiano, V., & Dash, J. (2021, 2021/01/01/). Land surface phenology as indicator of global terrestrial ecosystem dynamics: A systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing, 171*, 330-347. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2020.11.019

Congalton, R. G. (2001). Accuracy assessment and validation of remotely sensed and other spatial information. *International Journal of Wildland Fire, 10*(4), 321-328. https://doi.org/https://doi.org/10.1071/WF01031

ENVI. (2021). *Exelis Visual Information Solutions*.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data [Review]. *Remote Sensing of Environment, 186*, 64-87. https://doi.org/10.1016/j.rse.2016.08.013

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2015). VSURF: An R package for variable selection using Random Forests. *The R Journal 7/2*.

Ghosh, A., & Joshi, P. K. (2014, 2014/02/01/). A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation, 26*, 298-311. https://doi.org/https://doi.org/10.1016/j.jag.2013.08.011

Glenn, E. P., Huete, A. R., Nagler, P. L., & Nelson, S. G. (2008). Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape [Review]. *Sensors, 8*(4), 2136-2160. https://doi.org/10.3390/s8042136

Grabska, E., Hostert, P., Pflugmacher, D., & Ostapowicz, K. (2019). Forest stand species mapping using the sentinel-2 time series [Article]. *Remote Sensing, 11*(10), Article 1197. https://doi.org/10.3390/rs11101197

Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., & Atzberger, C. (2019). Optimal Input Features for Tree Species Classification in Central Europe Based on Multi-Temporal Sentinel-2 Data. *Remote Sensing, 11*(22), 2599. https://doi.org/10.3390/rs11222599

Immitzer, M., Vuolo, F., & Atzberger, C. (2016, 2016
2018-10-05). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing, 8*(3), 166. https://doi.org/10.3390/rs8030166

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008, 2008/10/15/). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment, 112*(10), 3833-3845. https://doi.org/10.1016/j.rse.2008.06.006

Laborte, A. G., Maunahan, A. A., & Hijmans, R. J. (2010). Spectral Signature Generalization and Expansion Can Improve the Accuracy of Satellite Image Classification. *Plos One, 5*(5), e10516. https://doi.org/10.1371/journal.pone.0010516

Land Information New Zealand (LINZ). (2020). *NZ 8m Digital Elevation Model (2012)*. Retrieved 18 December, 2020, from https://data.linz.govt.nz/layer/51768-nz-8m-digital-elevation-model-2012/

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news, 2*(3), 18-22.

Mallinis, G., Koutsias, N., Tsakiri-Strati, M., & Karteris, M. (2008, 2008/03/01/). Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site. *Isprs Journal of Photogrammetry and Remote Sensing, 63*(2), 237-250. https://doi.org/10.1016/j.isprsjprs.2007.08.007

Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016, 2016/12/15/). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment, 187*, 156-168. https://doi.org/10.1016/j.rse.2016.10.010

Persson, M., Lindberg, E., & Reese, H. (2018). Tree Species Classification with Multi-Temporal Sentinel-2 Data. *Remote Sensing, 10*(11), 1794. https://doi.org/10.3390/rs10111794

R Core Team. (2013). R: A language and environment for statistical computing.

Schuster, C., Förster, M., & Kleinschmit, B. (2012, 2012/09/10). Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing, 33*(17), 5583-5599. https://doi.org/10.1080/01431161.2012.666812

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019, 2019/11/15/). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications, 134*, 93-101. https://doi.org/10.1016/j.eswa.2019.05.028

Wang, B., Jia, K., Liang, S., Xie, X., Wei, X., Zhao, X., Yao, Y., & Zhang, X. (2018). Assessment of Sentinel-2 MSI Spectral Band Reflectances for Estimating Fractional Vegetation Cover. *Remote Sensing, 10*(12), 1927.

Ye, N., Morgenroth, J., Xu, C., & Chen, N. (2021, 2021/10/01/). Indigenous forest classification in New Zealand – A comparison of classifiers and sensors. *International Journal of Applied Earth Observation and Geoinformation, 102*, 102395. https://doi.org/10.1016/j.jag.2021.102395

Zhang, F., & Yang, X. (2020, 2020/12/15/). Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sensing of Environment, 251*, 112105. https://doi.org/10.1016/j.rse.2020.112105