Deforestation Mapping 2019 and 2020
Technical Report

Prepared for

Ministry for the Environment

26 July 2022

CITE REPORT AS:

New Zealand Deforestation Mapping 2019 & 2020 – Technical Report

Prepared by Lynker Analytics Consortium for the Ministry for the Environment

Table of Contents

List of Figures

List of Tables

Executive Summary

The Land Use Carbon Analysis System (LUCAS) is used for reporting on Land-Use Change and Forestry in New Zealand's annual national greenhouse gas inventory report submitted to the United Nations to meet international reporting obligations. From a greenhouse gas point-of-view, the most important land use changes are afforestation (new forest planting) and deforestation (forest removal and conversion to another land use).

The Lynker Analytics Consortium consisting of three companies; Lynker Analytics, Carbon Forest Services and UAV Mapping NZ Ltd was commissioned to survey and classify a total of 6,775 distinct areas of potential forest loss. The methodology used in the previous (2017/2018) deforestation mapping project was again used in this analysis.

Using satellite imagery, the Ministry for the Environment (the Ministry) had previously mapped distinct areas of forest loss, which occurred during 2019 and 2020. Added to this were 1,300 areas that lost forest in earlier years, for which current land use was unknown. These forest-loss 'targets' range in size from approximately 1 to 348 hectares and cover a total area of approximately 54,597 hectares.

Between January and June 2022, the Lynker Analytics Consortium conducted an aerial survey of all areas spanning every region of New Zealand using Cessna 172 aircraft flying at approximately 3,000 feet above ground level (AGL). This delivered 0.20m resolution vertical aerial photography in over 99% of targets. 62 targets were not acquired due to poor weather conditions or inaccessibility. High resolution satellite imagery from Maxar has been identified as suitable for these missed targets.

The imagery was georeferenced and then classified into land cover classes such as cutover, plantation seedlings, pasture, and mature native forest within a target using a Machine Learning (ML) model. From this, a geospatial data generalisation routine was applied to filter out small land cover groupings less than $320m^2$ followed by a multi-criteria iterative analysis to assign each area of forest loss a dominant land cover and replant status. The ML model was used as the primary monitoring system to detect deforestation and re-planting and flag those targets to the Ministry.

Of the 6,775 targets 1,874 were identified as not yet re-planted while 1,545 were classified as re-planted or partially re-planted. A total of 2,092 areas were found to be partially or fully deforested. In terms of total area the largest single land cover category was cutover (41%) followed by plantation seedlings (19%), natural regenerating forest (11%) and pasture (10%).

Overall, this approach facilitated land-use mapping in detail within target perimeters enabling greater alignment with the definitions of 'forest land', 'non-forest land', and 'deforestation' under the international and domestic rules.

The automated monitoring system proved reliable in detecting deforestation, re-planting and other land cover changes exceeding one hectare.

1 Scope and Methodology

The LUCAS programme uses multispectral satellite imagery to identify and map land use change across New Zealand to quantify the greenhouse gas emissions and removals associated with these changes. Over the summers of 2018/19, 2019/20 and 2020/21 Sentinel-2 satellite imagery was acquired and mosaicked into national snapshots of the country for each summer period. These snapshots were then compared to identify areas of forest loss, which occurred during 2019 and 2020.

The objective of the 2022 deforestation mapping project was to field check these areas of identified forest loss by acquiring high-resolution aerial photos of each of the areas, classifying the current land cover in these areas and ultimately determining which areas have been replanted.

The project consisted of two primary activity areas:

1. acquire evidence images for the full extent of all forest loss targets

2. produce a spatial layer of forest loss target boundaries attributed with the evidence image name and replanting-related attribution (e.g. “fully replanted”; “partially replanted”; or “not replanted”) based on the evidence image.

To produce the final forest loss spatial layer, we generated an intermediate layer which reports land cover at an 8m spatial resolution, referred to as super pixel, across each target. This was the final output from a deep learning land classification model fine-tuned and re-trained on the imagery acquired in the project. The base model for this was the model used in the 2017 and 2018 deforestation mapping which is described in this report.

Our solution encompassed five project phases run in a serial process to generate the evidence images and output data. The stages are depicted in Figure 1 and described at a high level below.

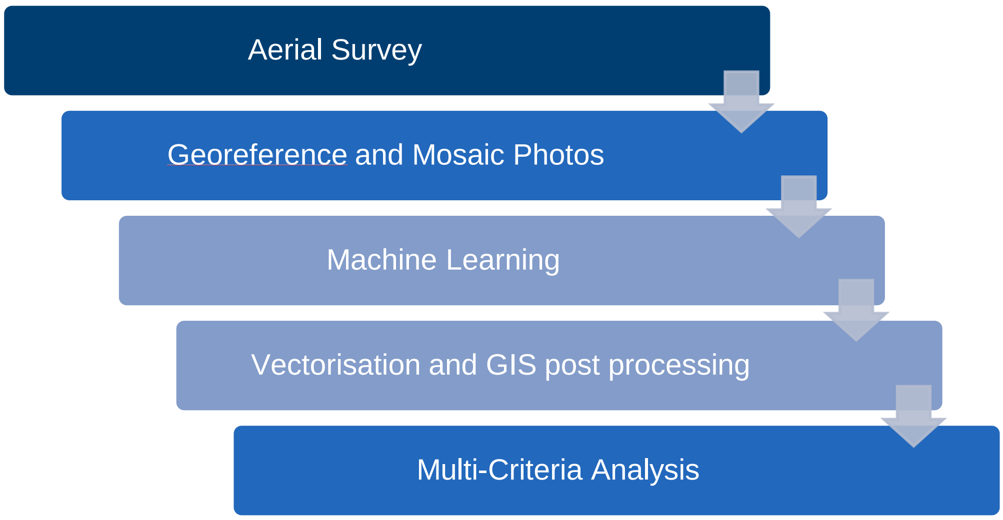


Figure 1. Solution workflow and processing pipeline

The aerial survey produced an individual photo mosaic for each deforestation target. Following image checks, these photo mosaics were then staged for input into a deep learning model. We used a patch-segmentation method to classify ground cover in patches of 64m^2 within every target. The land cover image was then vectorised, and filtering was applied to remove land cover areas less than 300 m^2 (considered noise) from the model output. This resulted in a land cover map for each deforestation target inclusive of up to 10 classes including plantation seedlings, cutover, mature exotic, mature native, and pasture.

Finally, a multi-criteria analysis was to the sub-area polygons to assess the ratio of land cover within each target. The result of this was the production of a spatial layer of forest loss target boundaries attributed with the evidence image name and replanting-related attribution.

2 Aerial Photography

The aerial survey was managed by UAV Mapping NZ Ltd using Cessna 172 aircraft (operated by Action Aviation) fitted with a full frame digital, nadir camera. Flight planning was carried out using the Aviatrix flight management system (FMS) and we used live-tracking GPS to fly a predetermined flight path.

To manage the national survey, image collection locations are loaded into a systematic and fuel-efficient order of capture. Once daily areas have been planned, a "classic traveler" least-distance analysis is run. This feeds into the on-board flight planner, while still allowing the pilot options to get into the best approach for each target.

The camera was a Canon 5D SR full-frame CMOS sensor fitted with a 24mm lens. The aircraft height was on average 3,300 feet above ground level (AGL). As the aircraft approached the area to be captured the camera triggered automatically and captured the image(s). A minimum of four nadir photos were captured over each target. The image mosaic footprint (orange box) for three areas is shown in Fig 2.

Figure 2. Image area for deforestation targets

Each photo was geotagged with WGS84 coordinates in decimal degrees, alongside detailed metadata including datetime, aircraft height, and pitch/roll/yaw. The camera acquired 3-band

RGB images which were then colour-balanced to settings that would optimise the ML process downstream. The aerial survey started on January 13, 2022 and completed on May 8, 2022. A summary of the aerial survey is provided in Table 1.

Metric	Quantity
Photos Captured (50mp 12bit raw images)	28,687
Acquired Targets	6,713
Flying Days	48
Flight Hours	198
Flight Distance (km)	23,371
Safety Issues	0

Table 1. Aerial Survey Summary

SimActive Correlator3D™ was used to undertake aerial triangulation of the images with positional data acquired across an entire flight used alongside existing orthophotography to tie photos together. The positional accuracy of the final 0.20m GSD imagery is approximately +/- 20m. All carbon emissions associated with the aircraft operations were offset on a carbon certified forestry project.

3 Machine Learning

The multi-scale convolutional neural network (CNN) model used in the previous (2017 and 2018) deforestation project was fine-tuned and used again for this project. The existing model was firstly re-trained using approximately 5,000 labels from the 2022 survey. Active Learning – a methodology that iteratively selects and presents image chips based on entropy for labelling was used to re-train the model. The selected image chips are then reviewed and annotated by human experts and added to a pool of labelled data for model retraining.   This max-entropy sampling method ensures only the most informative samples are reviewed and labelled by a human expert, leading to savings in human effort and processing time. This sampling method also mitigates the effects of sampling bias in the training set. Each labelled example (Fig 3) includes a context area of 64m x 64m and a detailed super pixel area of 8m x 8m (red box).

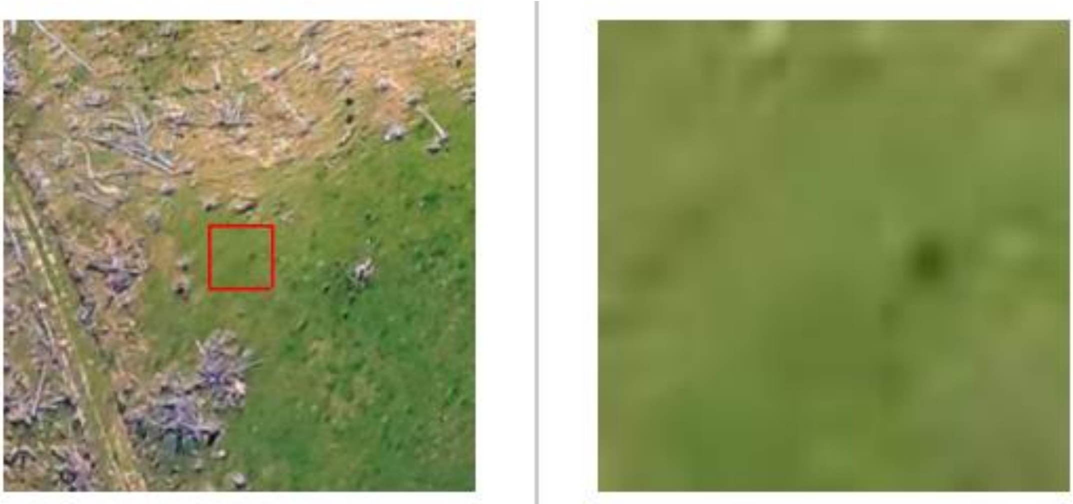


Figure 3. Training Annotations, plantation seedlings (top), pasture (bottom)

The neural network model uses a patch-segmentation method trained on the multi-scale image chip pairs. We used Google's Inception V3-based neural networks which are pre-trained on the ImageNet dataset.

These are well understood models and a good compromise between efficiency and performance. Transfer learning and fine-tuning was then applied to the neural networks, producing a fully connected classifier which combined the context view and a detailed view to classify $64m^2$ image chips.

The machine learning algorithm classifies all super pixels within the target area image into one of the following classes.

1. Built Forest (cut site, access tracks)
2. Built Other (pavement, buildings)
3. Crop
4. Cutover
5. Exotic Regenerating Forest (wildings)
6. Grass/Pasture
7. Horticulture
8. Mature Exotic Forest
9. Mature Native Forest
10. Natural Other (scree, riverbank)
11. Natural Regenerating Forest
12. Other
13. Plantation Seedlings
14. Shadow
15. Water

Table 2. Land cover categories

4 Model Accuracy

The following discussion examines the accuracy of the primary class sub-area land cover machine learning model predictions relative to validation data captured by our team. The validation data was not used to train the machine learning models but was captured at the same time and using the same process as the training data.

Table 3 presents several statistical parameters including:

- True Positive (TP) = the model correctly predicts the positive class
- False Positive (FP) = the model incorrectly predicts the positive class
- True Negative (TN) = the model correctly picks the negative class
- False Negative (FN) = the model incorrectly predicts the negative class
- Precision = TP / (TP + FP)
- Recall = TP / (TP + FN)
- F1 = 2 X ((Precision X Recall) / (Precision + Recall))

Precision, Recall and F1 scores provide additional measures of accuracy by class. Precision shows how precise/accurate the model is out of those predicted to be positive. Recall calculates how many of the actual positives the model capture through labeling it as True Positive. The F1 score is a useful measure of overall model accuracy as it represents a weighted average of Precision and Recall.

Class	Precision	Recall	F1	Samples
Cutover	0.68	0.75	0.71	114
Pasture	0.87	0.87	0.87	315
Mature Native Forest	0.88	0.82	0.85	245
Plantation Seedlings	0.69	0.85	0.76	125
Natural Regenerating Forest	0.47	0.49	0.48	88
Other	0.63	0.49	0.55	137
Accuracy			0.76	1024
Weighted average	0.76	0.76	0.76	1024

Table 3. Classification Statistics

Overall, the combined model accuracy is **0.76 (76%)** which is identical to the 2017/18 result. This is calculated by dividing the number of correct predictions by the number of total predictions. The weighted F1 score is also **0.76** which is a good result overall given the heterogeneity of these environments thus enabling the model to be used as an effective detection system for change. The subsequent filtering and multi-criteria analysis within the GIS provides further generalisation value to deliver a refined sub-area land cover map used to determine final re-plant status.

5 Multi-criteria Analysis

Following the machine learning phase, a GIS process is used to vectorise, eliminate very small land cover units from the inference results before a final multi-criteria analysis assigns a dominant land cover and re-plant status. These two stages are detailed below.

5.1 Vectorisation

The deep learning models produce multi-class raster images with a $64m^2$ resolution. These images span the entire image footprint which extends beyond the deforestation target polygon (Fig 2). Each raster is converted to a polygon feature class followed by an eliminate polygon process which eliminates a small polygon by merging it with the polygon from the surrounding features that it shares the longest boundary with. In this case polygons with an area size of less than 320 square metres were eliminated. This removed minor isolated land cover groupings which, while interesting, were not important to the final land cover determination at a target scale. The final feature class was then clipped back to the target boundary. The final polygon layer for each target holds the area of each class in square metres. This information is then used in the second stage analysis.

5.2 Multi-Criteria Analysis

Taking result polygons from the previous stage we applied a sequential decision process to classify the target polygon into dominant land cover categories broadly in line with the Emissions Trading Scheme - Geospatial Mapping Information Standard. Each target polygon is tagged with a final classification based on the criteria and decision sequence shown in Table 4. This sequential process is designed to output attributes in order of importance to the review process, deforested targets being of high interest followed by re-planted targets. Any targets not categorised are assigned to Unknown and will require later review by the Ministry. These are often complex and ambiguous areas with scrub cover, natural regenerating forest or inclusive of multiple classes less than a hectare in area.

STEP	CRITERIA - Exclude Shadow, Mature Native, Mature Exotic, Built Forest	DOMINANT LAND COVER
1	Built Other + Pasture + Crop + Horticulture > 80%	Fully deforested
2	Built Other + Pasture + Crop + Horticulture > 1 Ha	Partially deforested
3	Plantation Seedlings > 70%	Fully re-planted
4	Plantation Seedlings > 1 Ha OR > 30%	Partially re-planted
5	Exotic Regen OR Cutover > 1 Ha or > 30%	Not re-planted
6	(Re)-include Mature Exotic > 1 Ha	Still Forest
7	Natural Damage + Natural Other > 1 Ha	Natural Adverse Event
8	All other	Unknown

Table 4. Multi-criteria decision process. Note at step 6 the class "Mature Exotic" which was previously excluded from steps 1-5 is now evaluated.

6 Results

The final breakdown of all targets following the complete analysis is shown in Figure 4.

Of the 6,775 targets 1,874 were identified as not yet re-planted while 1,545 were classified as re-planted or partially re-planted. A total of 2,092 areas were found to be partially or fully deforested. The unknown category contains mostly small areas that don't meet any of the previous qualifying criteria. These typically have no plantation seedlings and less than a hectare of any single land use change class such as pasture, crop or built/urban.

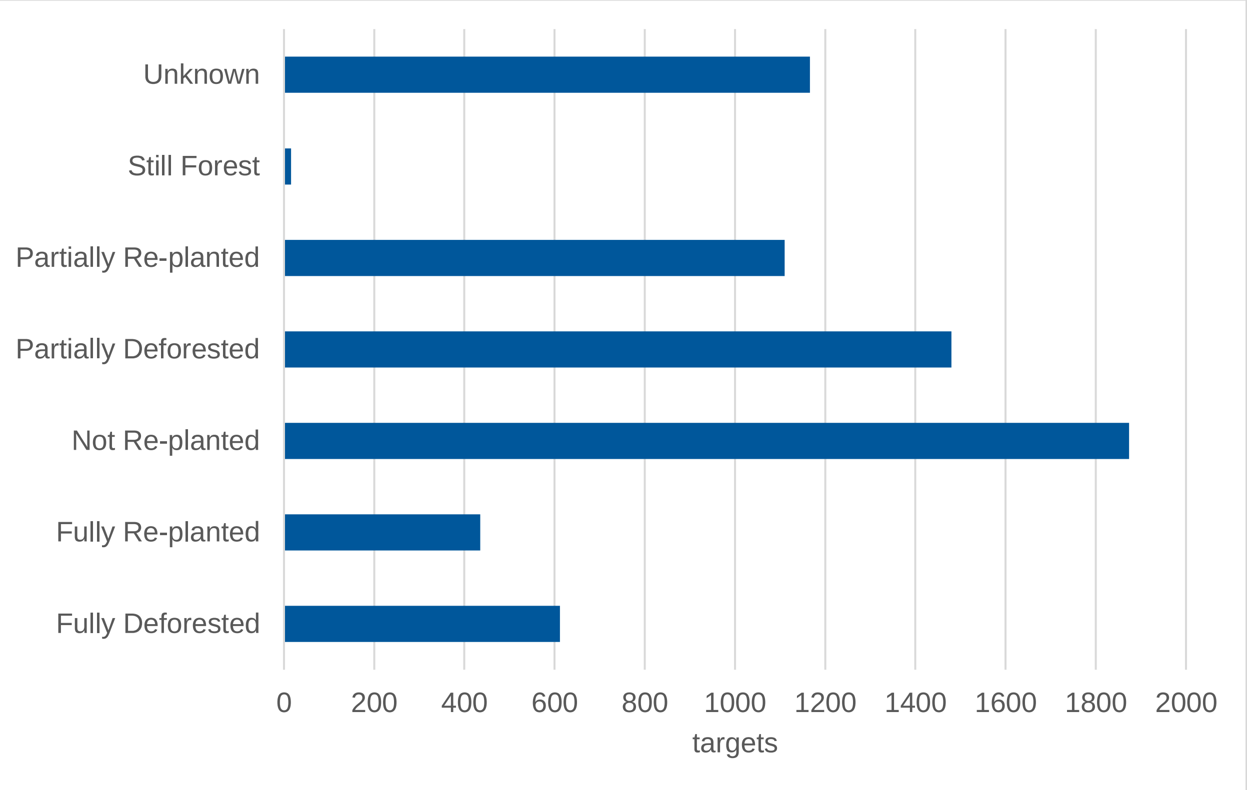


Figure 4. Final categorisation of the 6,775 deforestation targets

The breakdown of total area across all targets is shown in Figure 5. This methodology allows a more detailed analysis of land cover and land use across the country or by region as well as at a target level.

The largest single land cover category was cutover (21,905 ha or 41%) followed by plantation seedlings (19%), natural regenerating forest (11%) and pasture (10%).

Just under 4,000 hectares of mature native forest was mapped while 3,000 hectares was identified as containing built forest (harvest tracks, cut sites etc.).

Over 20,000 hectares of cutover was mapped across all areas indicating that many sites remain in a post-harvest state which aligns with the relatively large number of sites classified as "Not re-planted".

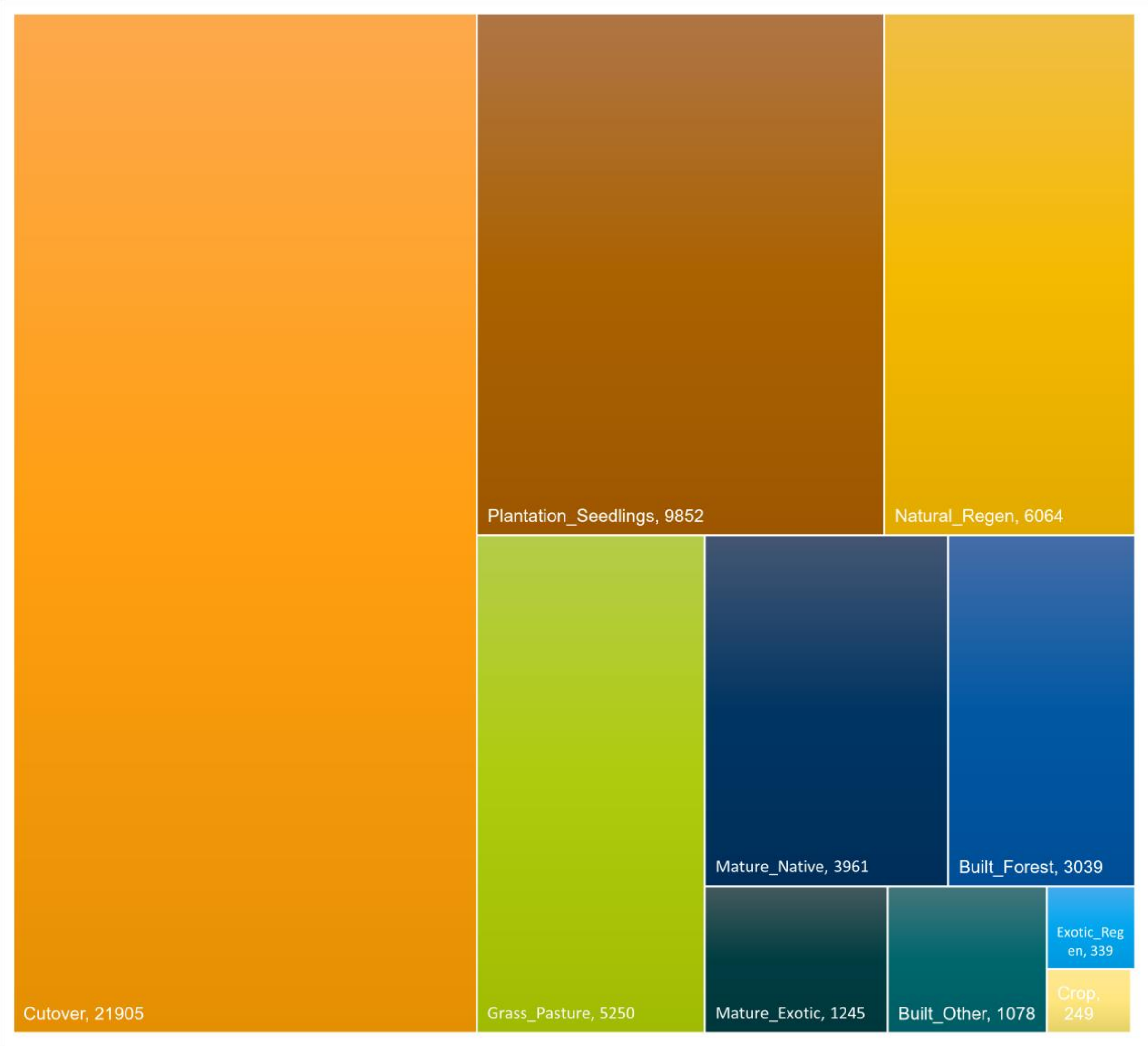


Figure 5. Land cover (hectares) across all targets

7 Deliverables

The deliverables include the following:

A. an updated Target Layer (polygon feature class) where each target has been updated with a dominant land cover classification and replanting status based on the method described in this report

B. for every Target, the applicable evidence image(s) are provided and named to match the ORIGINAL_RECON_IMAGE_NAME attribute supplied in the Updated Target Layer ("Evidence Imagery")

C. a file with image identifiers to Maxar satellite imagery suitable to visually identify the 62 targets not captured with aerial photography

D. a file geodatabase containing:

   a. a polygon feature class of footprints and metadata for all aerial imagery supplied.
   b. a point feature class of locations and metadata for all terrestrial imagery supplied.

E. a land cover layer delineating Target sub-area and attributing each sub-area with its land cover extent

F. Python code including machine learning model and model parameters

G. a final report (this document)

8 Key Learnings

The following list of key technical findings and lessons learned are provided for future reference. This should be reviewed in conjunction with the findings from the previous deforestation report completed for the 2017 and 2018 period.

1. **Imagery and rectification**

The image resolution was slightly higher in this project than in 2020 at 0.20m. This was to get below cloud cover on more days. Aerial triangulation of the images was achieved using SimActive Correlator3D™ with positional data acquired across an entire flight used alongside existing orthophotography to tie photos together. This was an improvement over the manual placement used in the previous survey.

The 2022 summer was very cloudy with ex-Tropical Cyclones Dovi, Eva, Fili, and the cloud and turbulent weather conditions of this La Niña season. Advancing collection into December in order to take advantage of finer weather would be beneficial.

2. **Machine Learning**

The multi-scale CNN models used in the previous survey were successfully adapted for this imagery set. A further 5,000 training annotations were acquired regionally and seasonally to fine tune the model. The baseline model was fine-tuned with emphasis on separating Exotic regen from Plantation Seedlings, augmentation of Plantation Seedlings with spot spray examples and examples with cutover and scrub distributed amongst young seedlings. The existing training pool was curated and refined to improve all categories with respect to the target imagery. A similar overall accuracy was achieved providing confidence that this now a proven and repeatable process.

3. **GIS Post-processing**

The two-stage post-processing approach worked well. The final multi-criteria analysis was adjusted based on feedback from the Ministry to those targets that may contain land use change of at least 1 hectare. Minor changes were made to how wilding pines (exotic regeneration) was handled in particular.

Appendix 1

Example Classification Results

Examples of the sub-area classification are provided below. Note the inference result is generated at an 8m x 8m spatial resolution. In each case the aerial photograph is shown alongside the classified polygon layer.

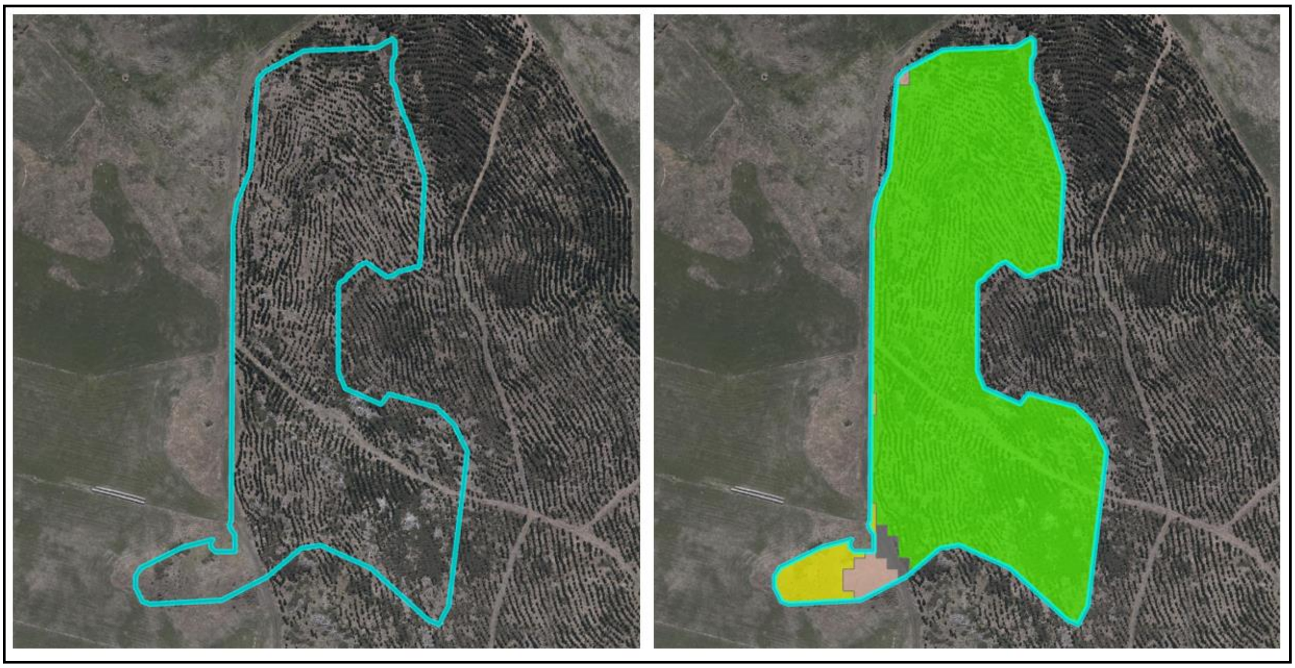
The polygon layer is the final delivered sub-area data set which has undergone GIS post-processing.

The classification legend is shown here. For each target we have additionally provided a summary of the following:

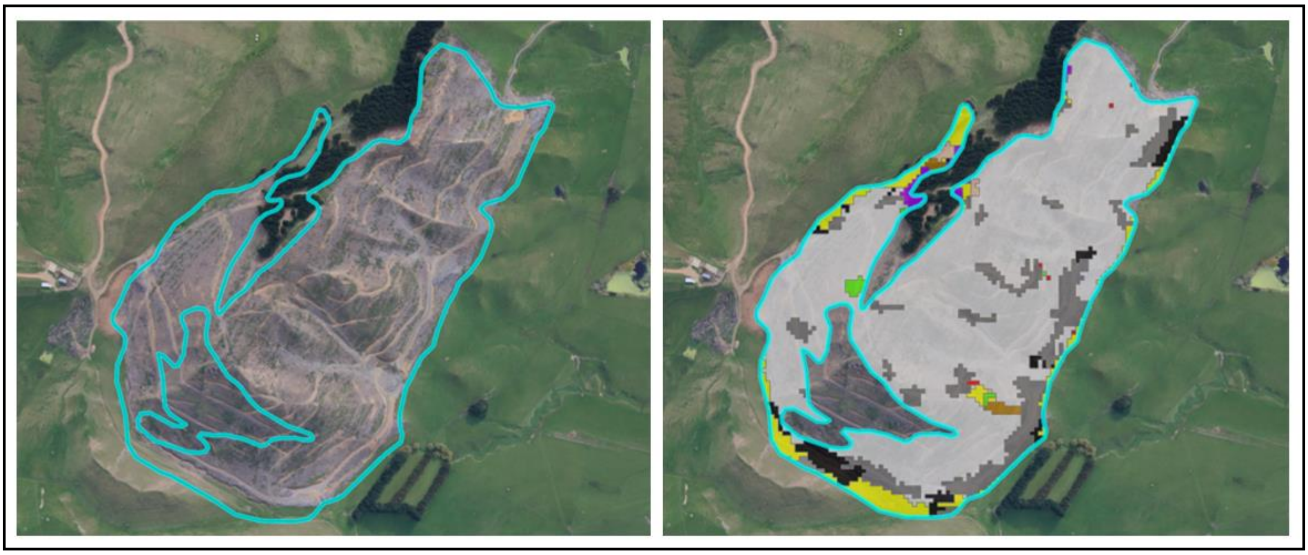
- target area in hectares
- region
- replant status

LKR_2020_MAN_0977 – Manawatu; 6.36 Ha; Fully Replanted

LKR_2020_TAR_0239 – Taranaki; 5.02 Ha; Fully Re-planted



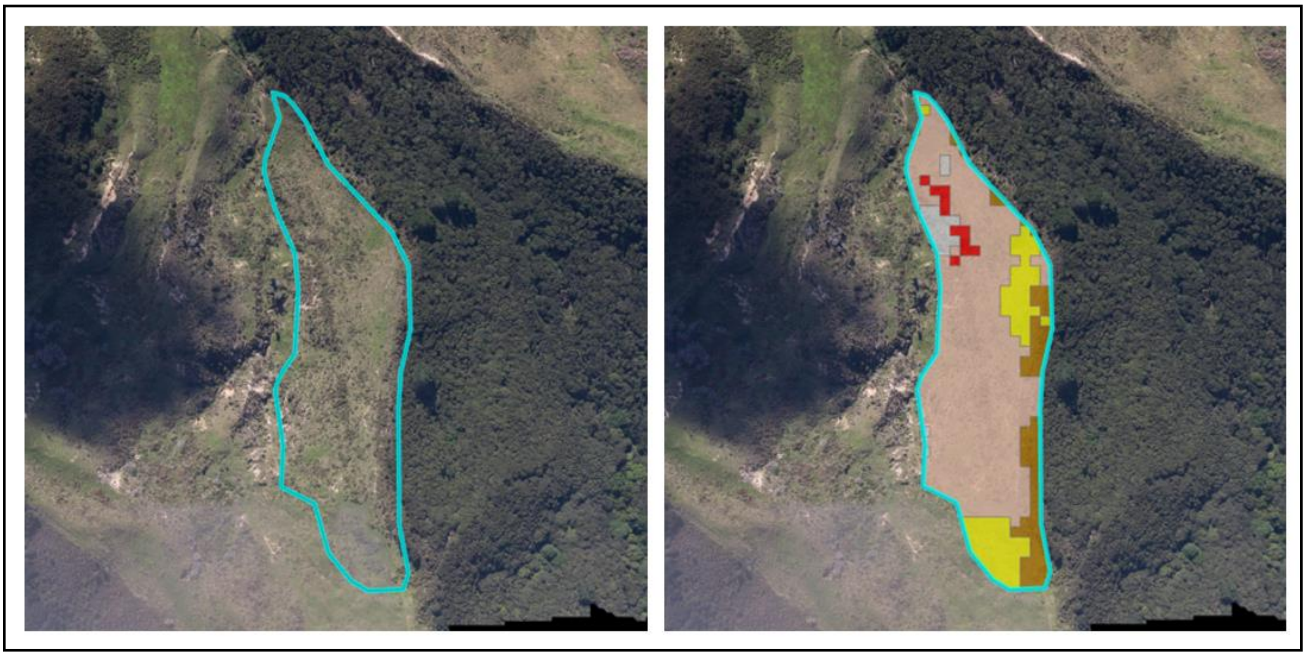
LKR_2020_HWK_0367- Hawkes Bay, 34.92 Ha; Not re-planted



LKR_2020_TAS_0635, Tasman, 1.65Ha, Fully Deforested

LKR_2020_WEST_0452, West Coast; 1.57Ha, Fully Deforested

LKR-2020_BOP_0297, Bay of Plenty, 1.83 Ha, Unknown



Appendix 2

Machine Learning Code

1. The project code was copied to the Ministry's GutHub repository.

2. The repository consists of

    a. Python code and a neural network model weights file.

3. The python script will run the ML land classification inference over RGB aerial imagery to produce a raw land cover classified raster.

4. The models are trained on vertical aerial photography with a spatial resolution of 0.20m.

5. The repository contains a requirements.txt file that lists the python libraries required to run the code and a README.md file that describes how to run the ML inference.

Lynker Analytics Ltd
Suite 6, Level 2
10 Hutt Road
Petone
Lower Hutt 5012
New Zealand

Carbon Forest Services
60 Cuba St
Wellington 6011
New Zealand

UAV Mapping NZ Ltd
14 Smiths Rd
RD-1
Ohakune 4691
New Zealand