Transforming tree phenotyping

SCION
FORESTS ▪ PRODUCTS ▪ INNOVATION

FOREST GROWERS RESEARCH

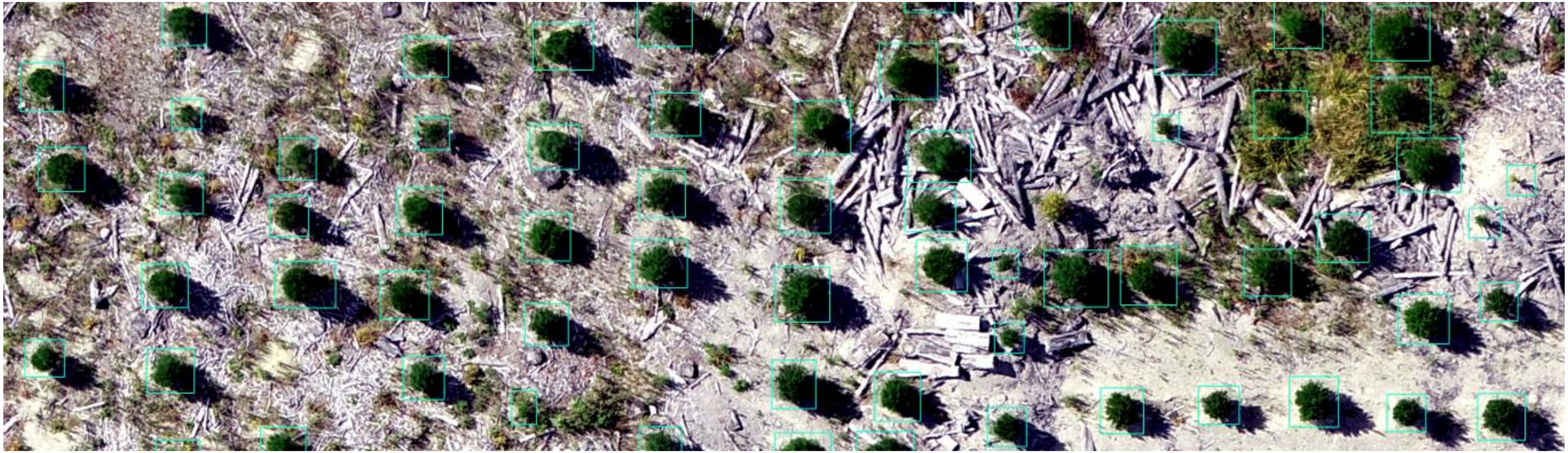
From shared datasets to fake datasets:
an update on deep learning work at Scion

Nicolò Camarretta

Remote Sensing cluster group - 13 November 2023

Unlocking AI through Shared Datasets:
Dataset partnerships for AI in forestry

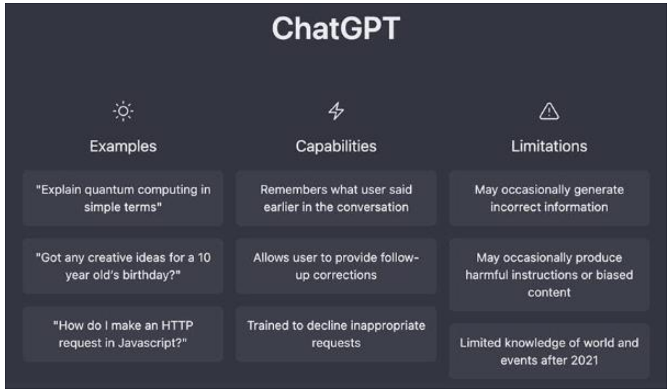
Nicolò Camarretta, Melanie Palmer, Ben Steer, Robin Hartley, Grant Pearse

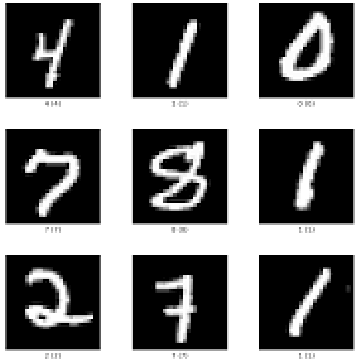


Background and motivation

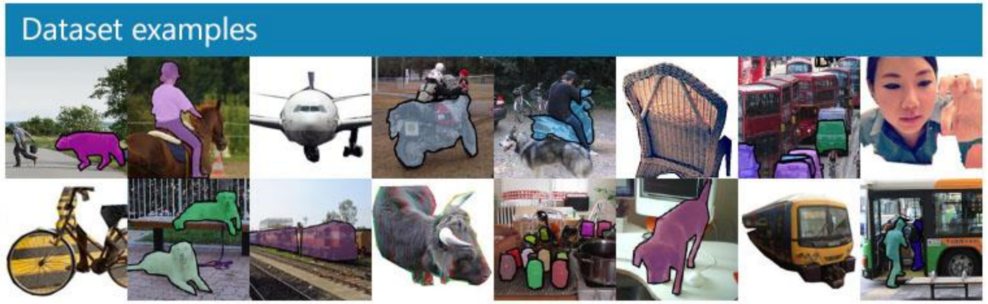
<u>Rapid</u> acceleration in ‘deep learning’ (AI):

- Computer vision largely solved
- Works well on geospatial imagery too: UAV, aerial & satellite
- General rule: If you can identify it, an algorithm can too …with a large enough dataset (and the right model)





MNIST: 70,000 digits



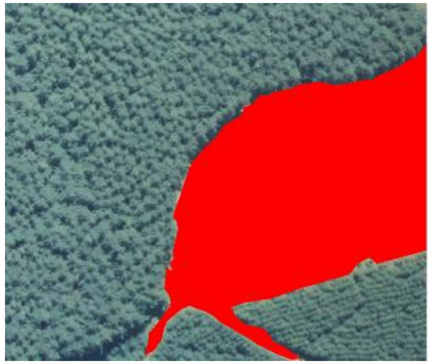
MS COCO: 164K in 80 categories

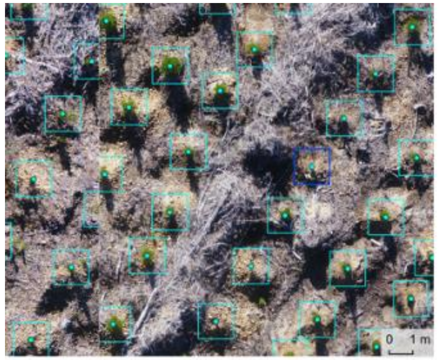
Imagenet: 1.4 M images in 1000 classes

Deep learning – forestry applications

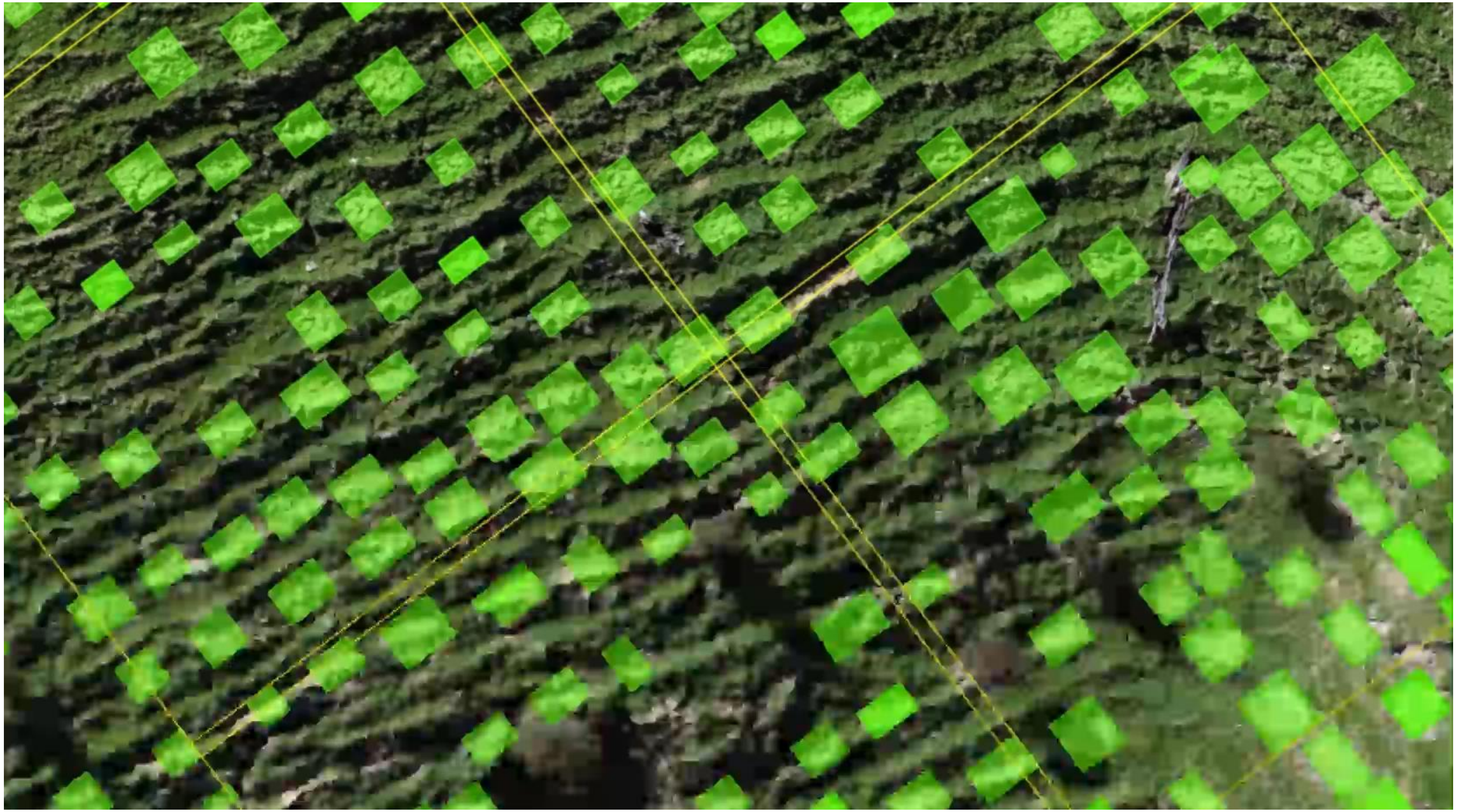
- Forestry companies are rapidly waking up to the potential for geospatial AI
- Scion – AI research programme
- Deep learning + UAV imagery = ♡
  - Seedling detection
  - Stand boundaries
  - Tree counting (post thinning)
  - Cutover mapping
  - Species ID
  - Canopy segmentation
  - 3D deep learning (stem, branch, etc) – LiDAR application

esri



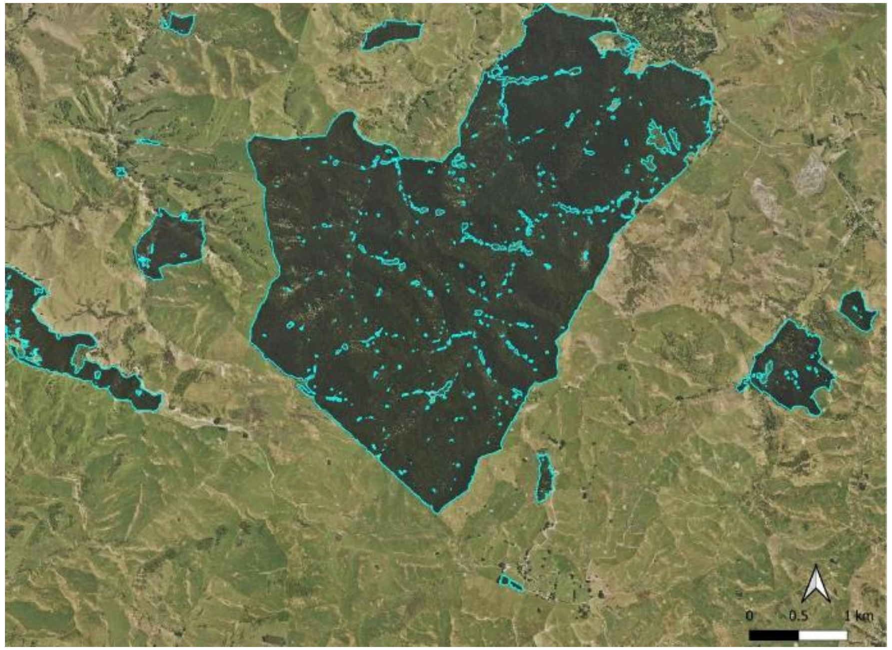


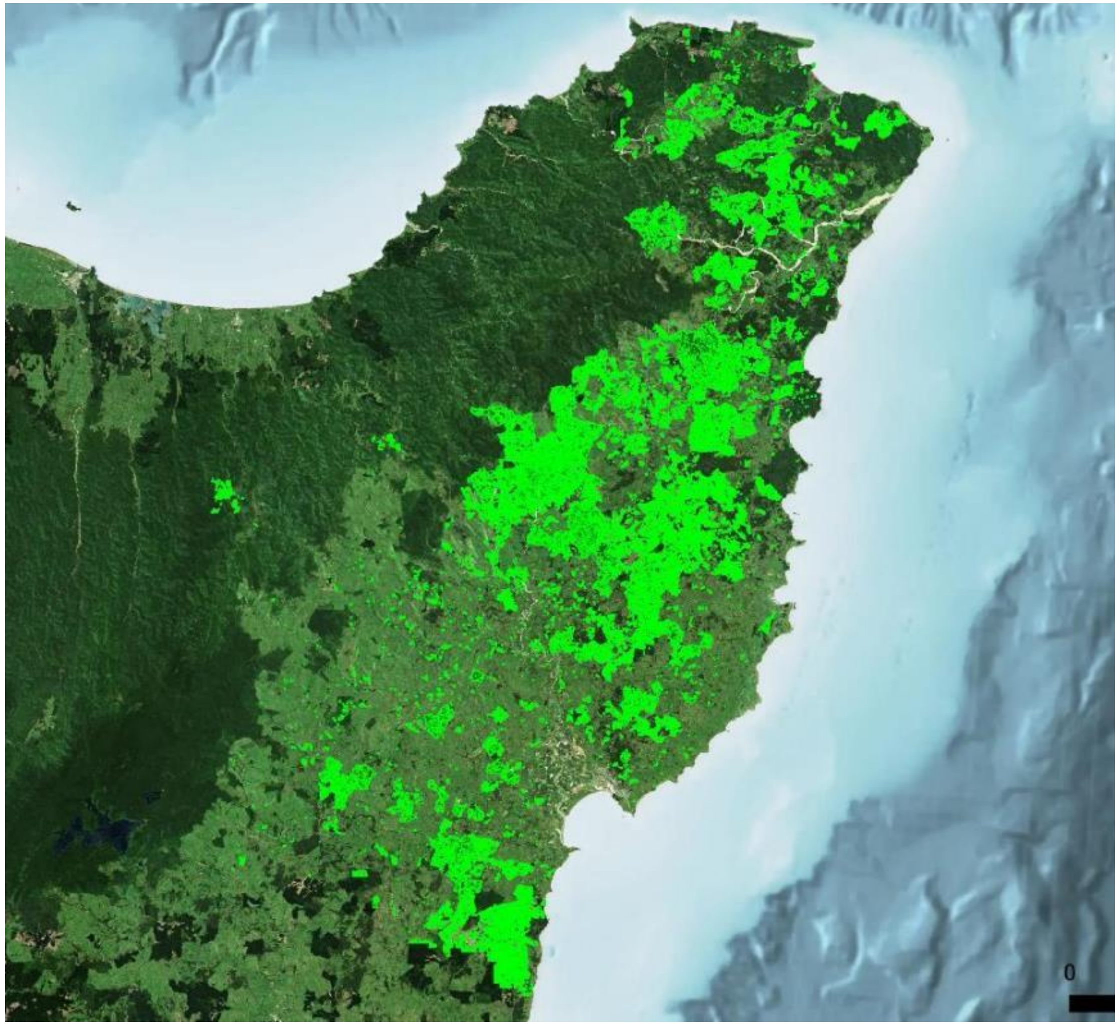
Deep learning – seedling detection



Deep Learning – stand boundary mapping

- Simple RGB imagery from a fixed wing aircraft is an important data source
- RGB used as input to deep learning to detect and delineate forest boundaries
- Recently used to detect damage from cyclone Gabrielle

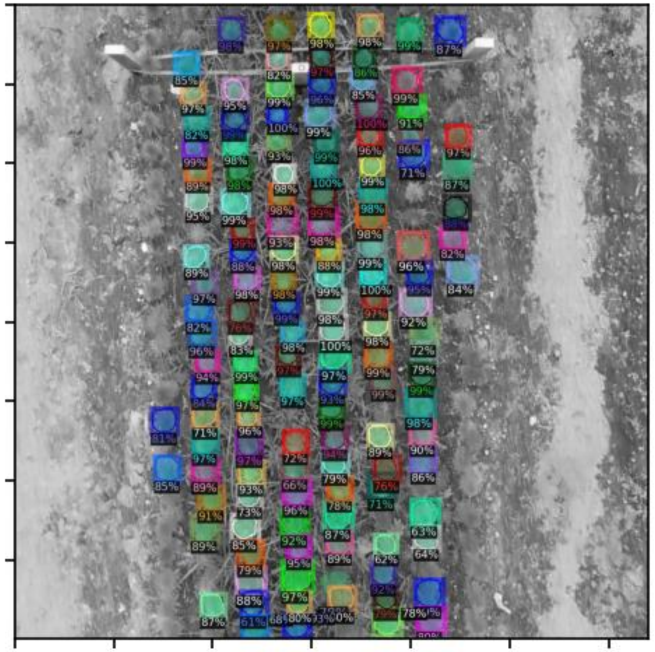




https://www.smartforest.cloud/

Deep learning – nursery phenotyping

- Early-crop phenotyping
- Deep learning detection and segmentation
- 20 images = 86% detection accuracy on test set



Deep Learning – Challenges

Barriers to uptake

- New skill, new processes, new concepts
- Terminology, jargon
- Rethink business practices
- Hard to get started
- The **<u>datasets</u>**

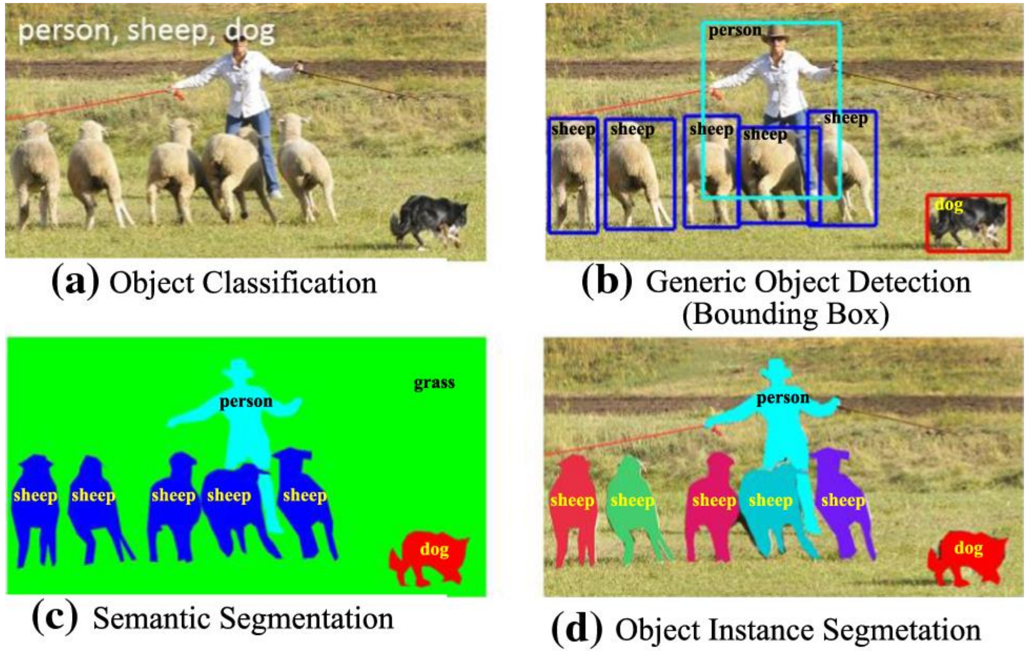


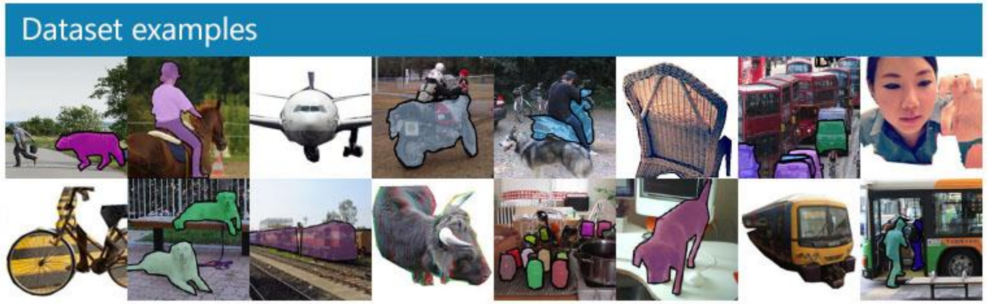
Image: Liu et al. 2020 https://doi.org/10.1007/s11263-019-01247-4

Dataset partnerships for AI in forestry

Why focus on datasets?

> *"[T]he "dirty secret" of artificial intelligence is that getting the software to work well in the real world requires a large amount of high-quality data."*
>
> - Alexander Wang, Founder & CEO Scale AI in an interview with *Fortune*

ADE20K: 25,000 densely annotated images



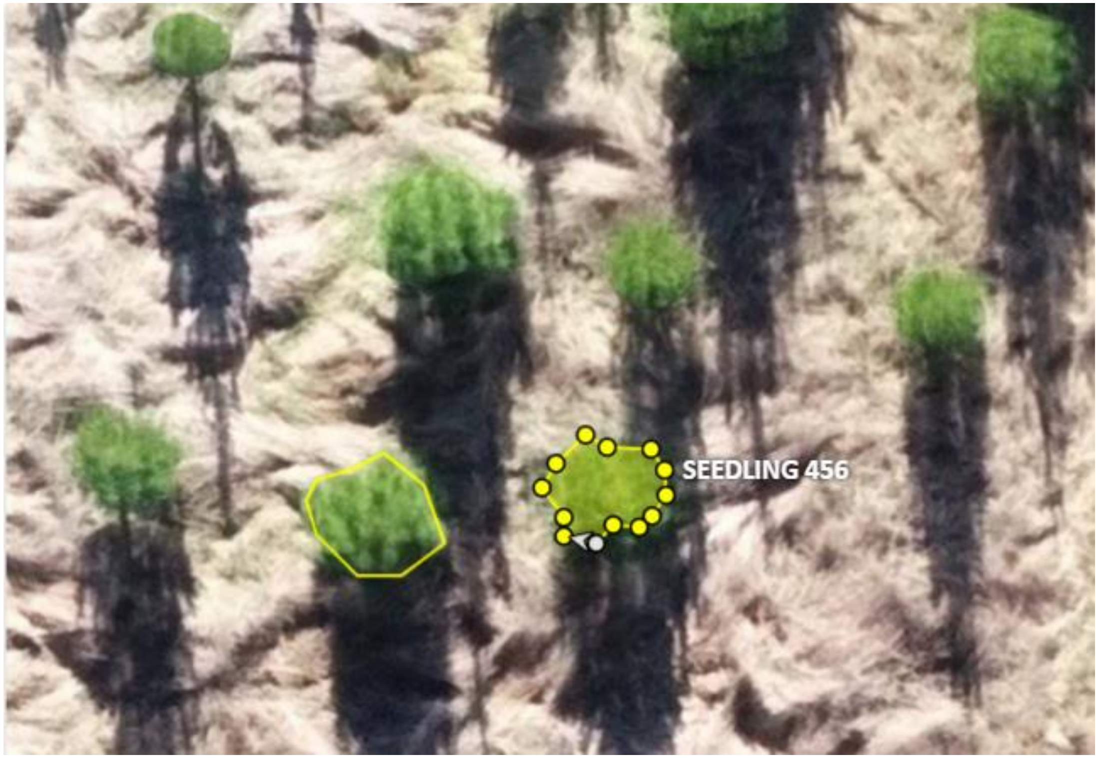
MS COCO: 164K in 80 categories

ImageNet 1.4 M images in 1000 classes

Dataset partnerships for AI in forestry

The dataset is often the hardest part

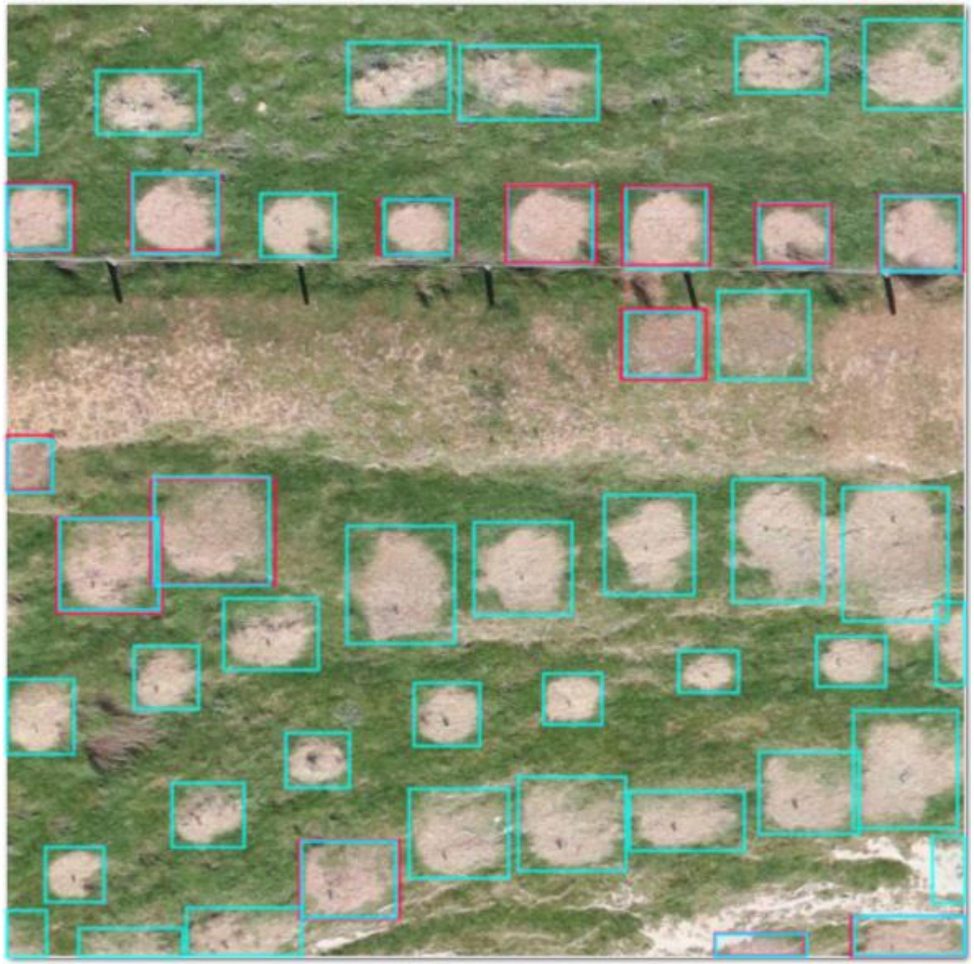
- Data acquisition is costly (UAV, plane, satellite)
- Data labelling can be **very** costly
- **Significant** time investment
- <u>**Quality is key**</u>
- It is an investment
- but…



Dataset partnerships for AI in forestry

Reasons to collaborate:

- Build large, diverse, high-quality datasets
  key to generalization
- Rapid dataset creation
- Efficient use of time and data
  - Focus data labelling effort on harder areas
  - Target collection of new data
- Scion's knowledge sharing: **get it right, the first time**
- Reduced costs (using existing data – imagery & GIS)



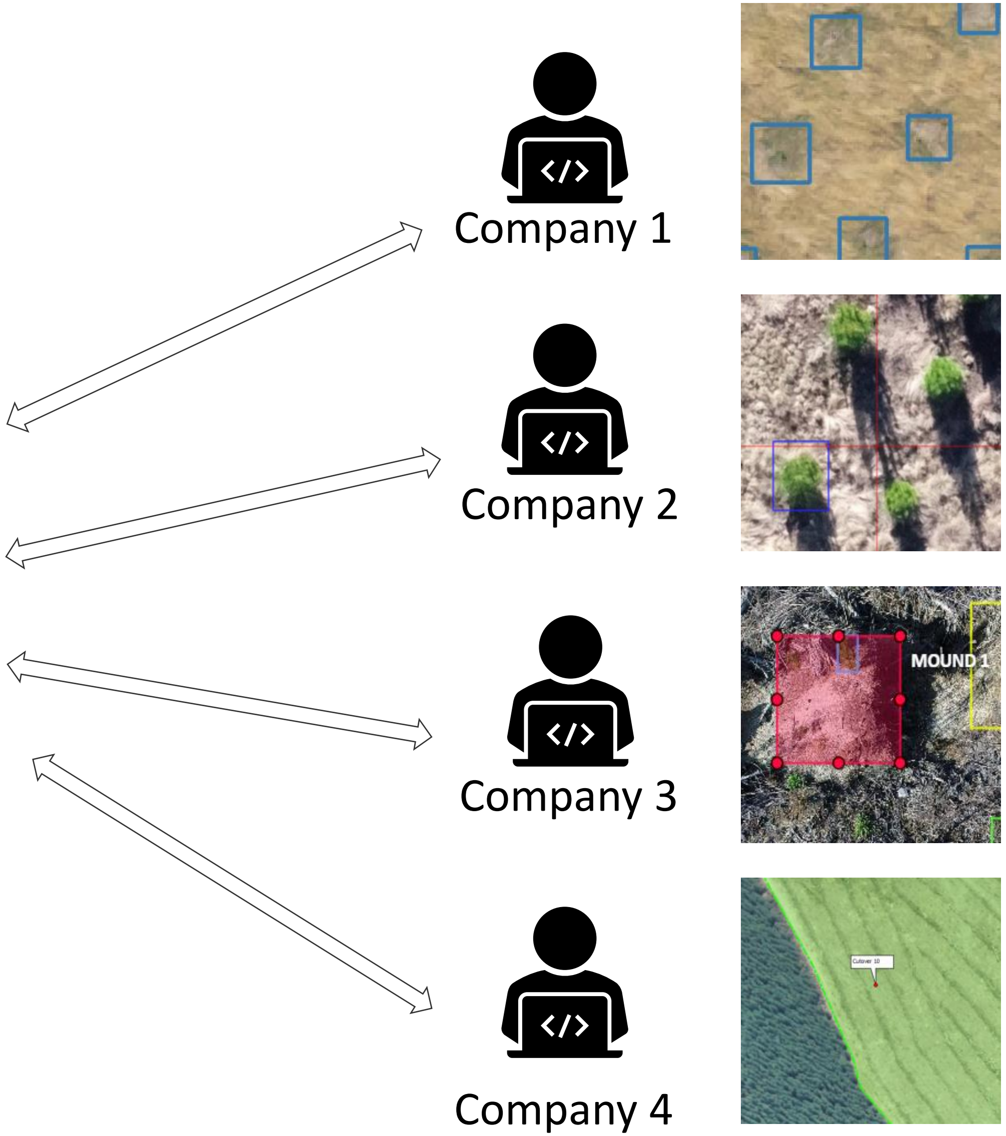
Dataset partnerships for AI in forestry

Dataset diversity is a good thing

Dataset partnerships for AI in forestry

Why use Zenodo?

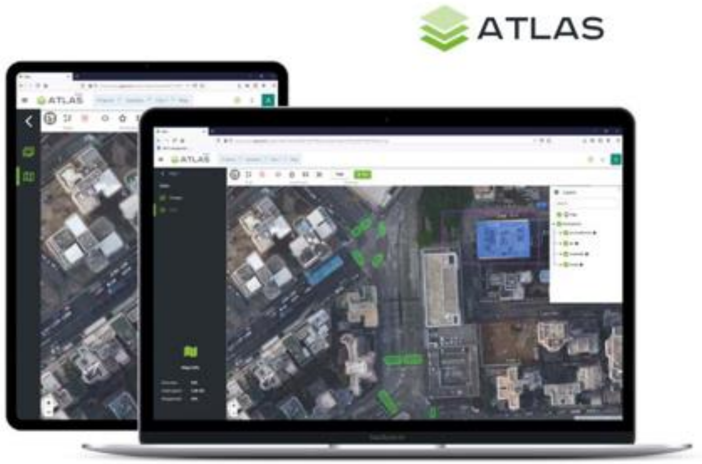
- **Safe** — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citeable** — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- **No waiting time** — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
- **Open or closed** — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
- **Versioning** — Easily update your dataset with our versioning feature.
- **GitHub integration** — Easily preserve your GitHub repository in Zenodo.
- **Usage statistics** — All uploads display standards compliant usage statistics

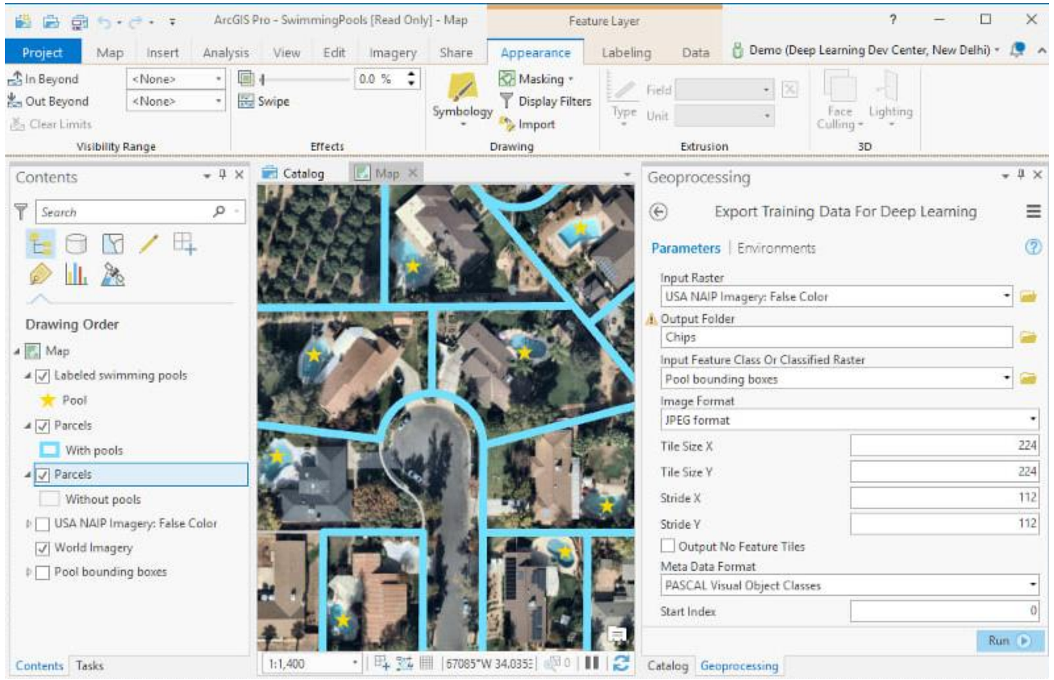


Dataset partnerships for AI in forestry

Why deliver the dataset?

- Having the dataset lets you train your own model
    - Continually improve and update
    - Power to train your own model
    - Combine data with commercial tools
- Value-add over duplicate dataset building





Why deliver the model?

- Becoming easier to deploy AI models
- Accessible tools (*e.g.*, ArcGIS Pro)
- Build familiarity with AI/ML in forestry

We are looking for partners

Please, get in contact!

Transforming tree phenotyping

A pipeline for generating high-fidelity synthetic point clouds for use in forest phenotyping

Scion Team:
Grant Pearse, Celine Mercier, Tancred Frickey, Grant Evans, Sadeepa Jayathunga, Robin Hartley, Elizaveta Graevskaya.

University of Sydney Team:
Ahalya Ravendran and Mitch Bryson

Our Goal

Develop a state-of-the-art phenotyping programme for radiata pine.

- Decision support system matching genotype to current and **future** climate
- Drought tolerance, disease resistance and carbon storage
- In-situ phenotyping pipeline that takes advantage of existing breeding trials
    - Partnership with Radiata Pine Breeding Company

RPBC®
Radiata Pine Breeding Co

Transforming Tree Phenotyping

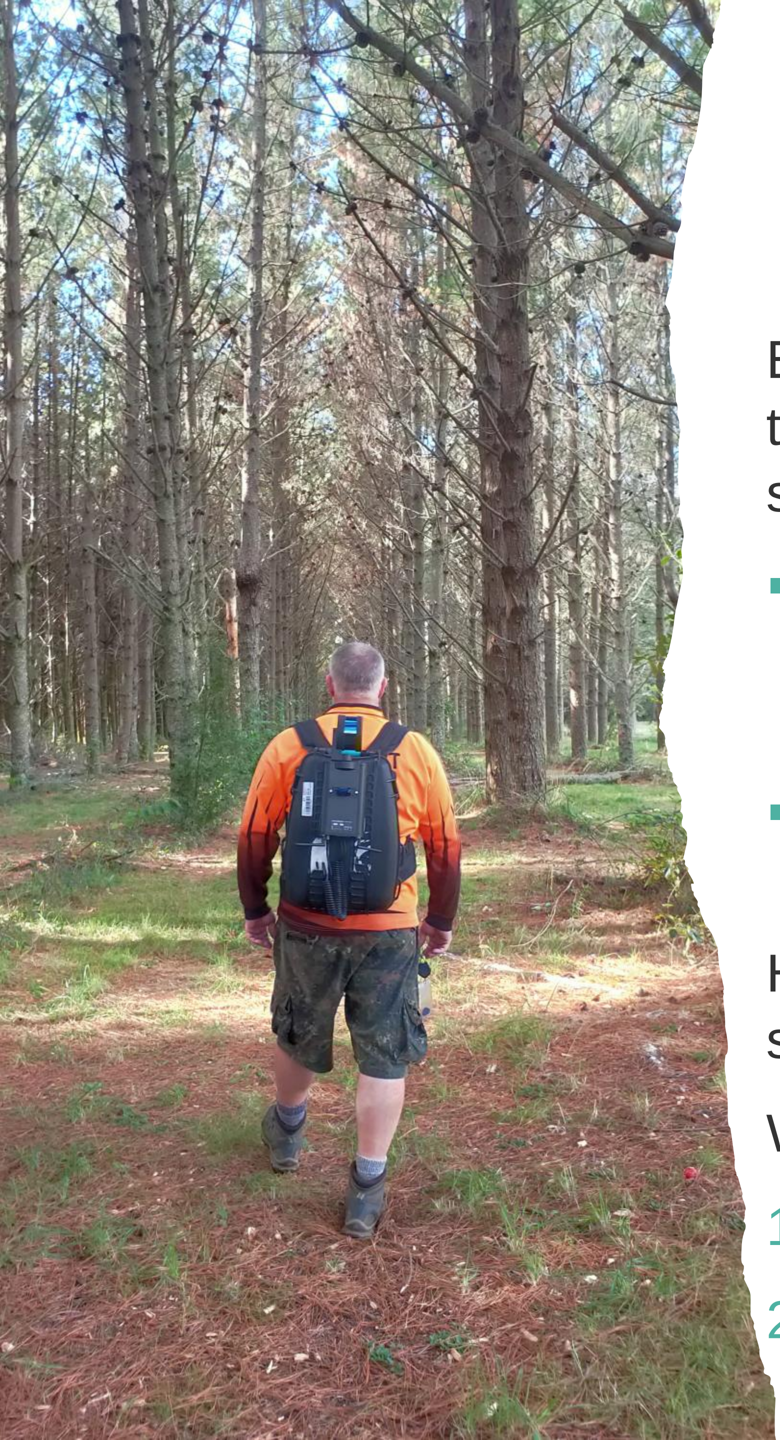
MBIE funded programme - NZ$9 million over 5 years

Programme lead: Michael Watt

1. Genetics: GxE
2. Cultural and native species phenotyping (Lania Holt)
3. Hyperspectral and thermal analysis for phenotyping (Michael Watt)
4. **Advanced Single Tree Characterisation (Robin Hartley)**

MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

Transforming tree phenotyping | scion



4. Advanced Single Tree Characterisation

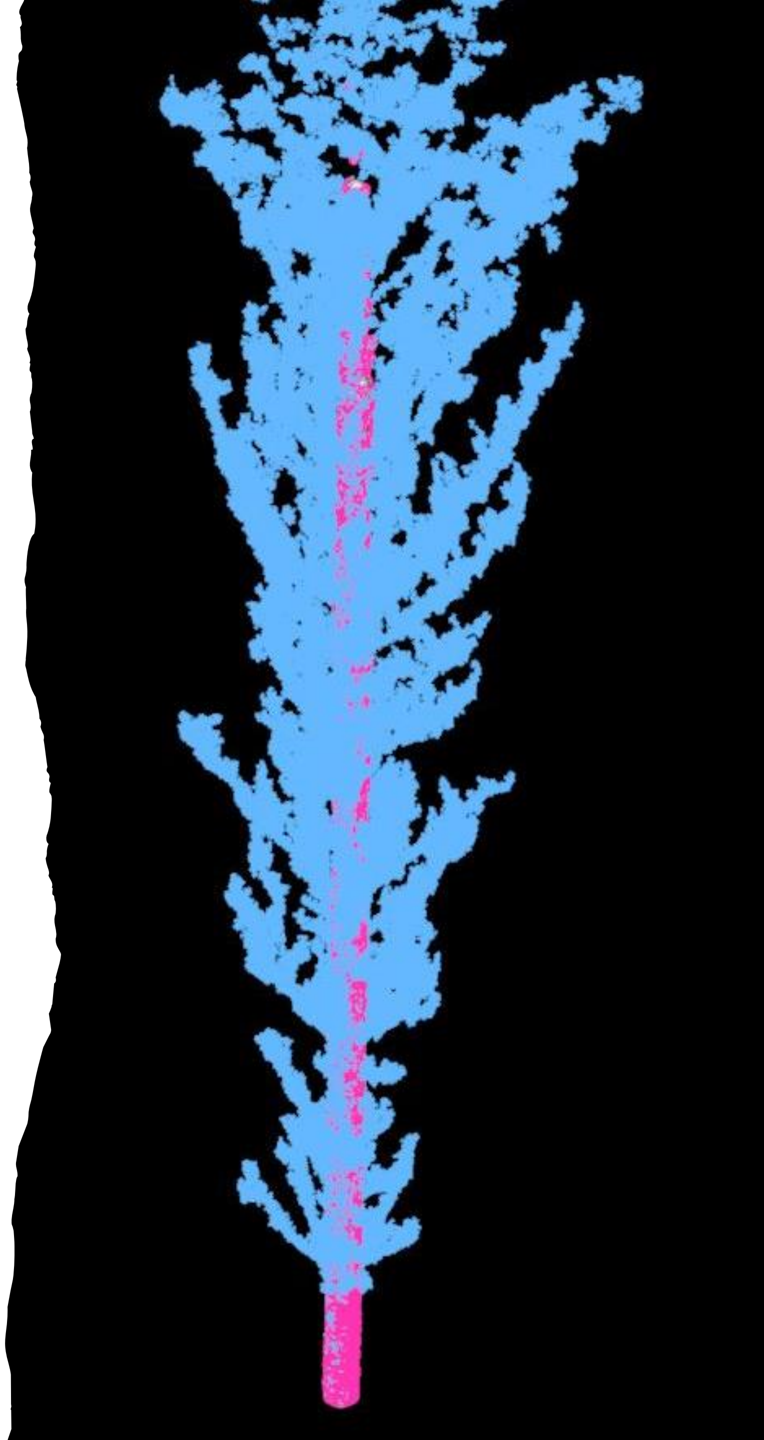
Extract structural attributes from breeding trials and mature stands using laser scanning.

- Structural selection criteria e.g., branch size, inter-whorl distance, branch angle (damage susceptibility).
- Carbon and stem volume are also important traits

How to extract these structural attributes at scale from point clouds?

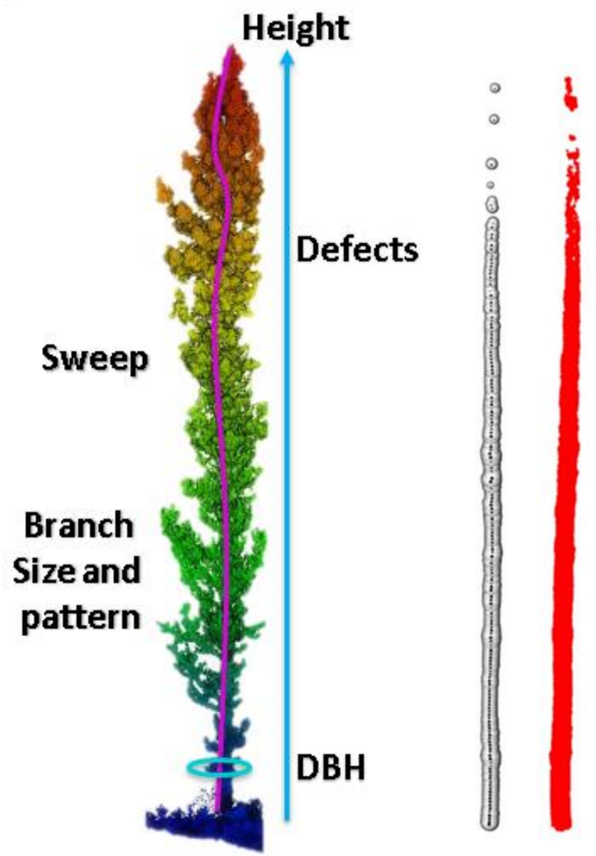
We use a two-pronged approach:

1. Quantitative Structural Models (QSM)
2. 3D deep learning on the point cloud



QSM for tree characterisation from 3D data

- Quantitative structure models (QSM)
  - Rule-based point cloud processing
  - Extract and reconstruct stem, branches etc.
  - Excellent results but one-by-one
- Scale and level of detail are **big** challenges
- Getting good results but it is a highly parameterized workflow



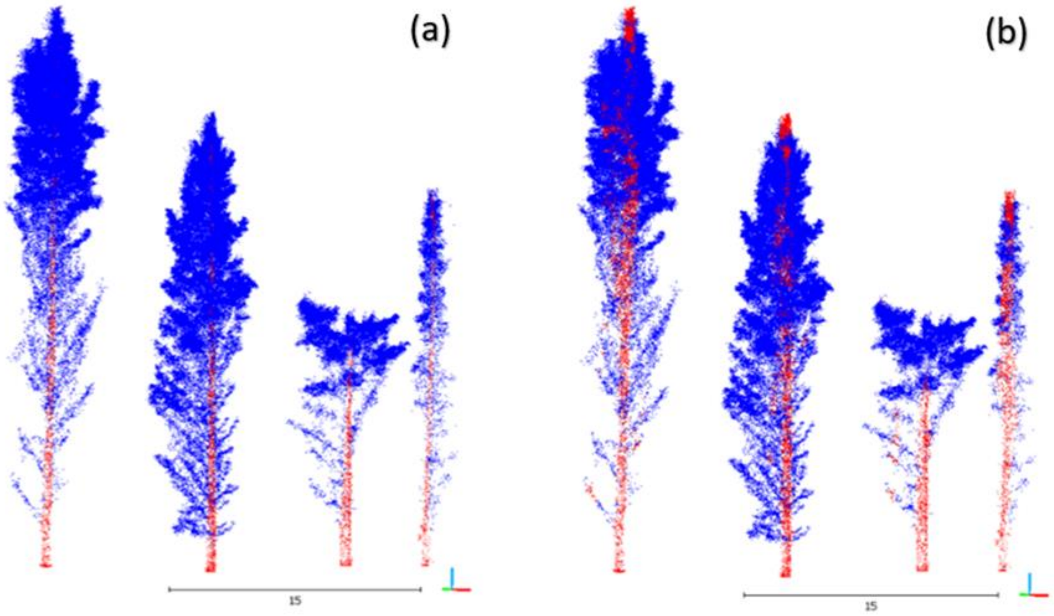
Hartley, R., Jayathunga, S., Massam, P., Davidson, S., De Silva, D., Estarija, H., Wuraola, A., Pearse, G. (2021). *Capture and extraction of phenotypic traits from novel high-density point clouds.* Resilient Forests Tech Note

3D Deep learning for tree characterisation

- FORInstance Dataset: Stefano Puliti - NIBIO
  - Benchmark dataset for ML classification of point clouds
  - Boreal to Tropical forest types
  - Manual annotation of thousands of trees
  - Hard work, expensive and some uncertainty in the labelling

Is there another way?

- Synthetic data
- Common in other domains



a) Ground truth b) Prediction

Hartley, R., Jayathunga, S., Massam, P., Davidson, S., De Silva, D., Estarija, H., Wuraola, A., Pearse, G. (2021). *Capture and extraction of phenotypic traits from novel high-density point clouds.* Resilient Forests Tech Note

Transforming tree phenotyping | scion

Synthetic data generation

blender

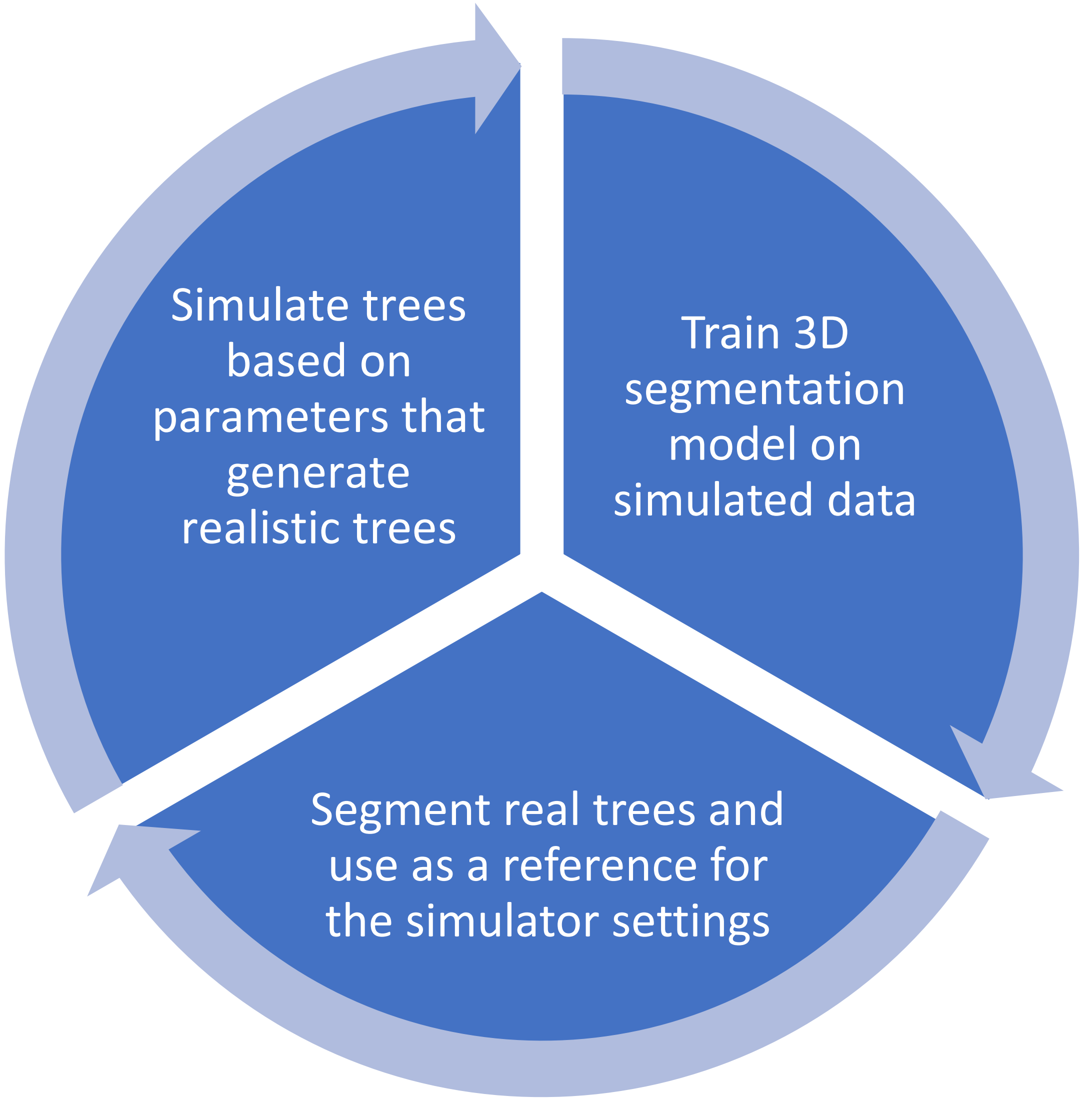
HELIOS++

Transforming tree phenotyping

scion

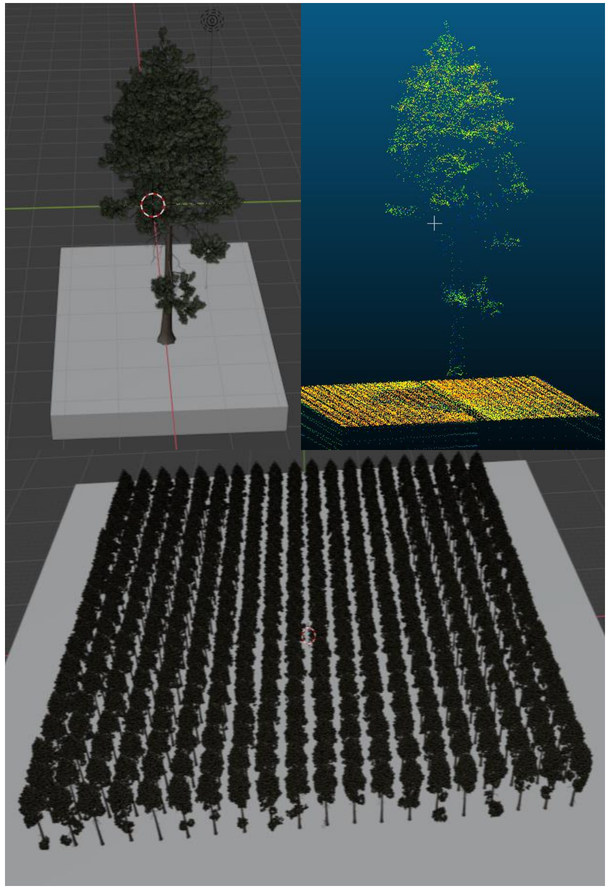
High-fidelity pipeline

- “Easy” to make nice-looking fake trees but they are not rooted in biological simulators
- Work with Singapore A*STAR Institute
- Closing the loop is hard as these trees are for video games and the parameters are only loosely connected to biology
- Lots of interaction between settings
- Current approach is to train a model to predict the parameters needed to make synthetic trees that look like real trees
- Add some jitter to the settings for variation

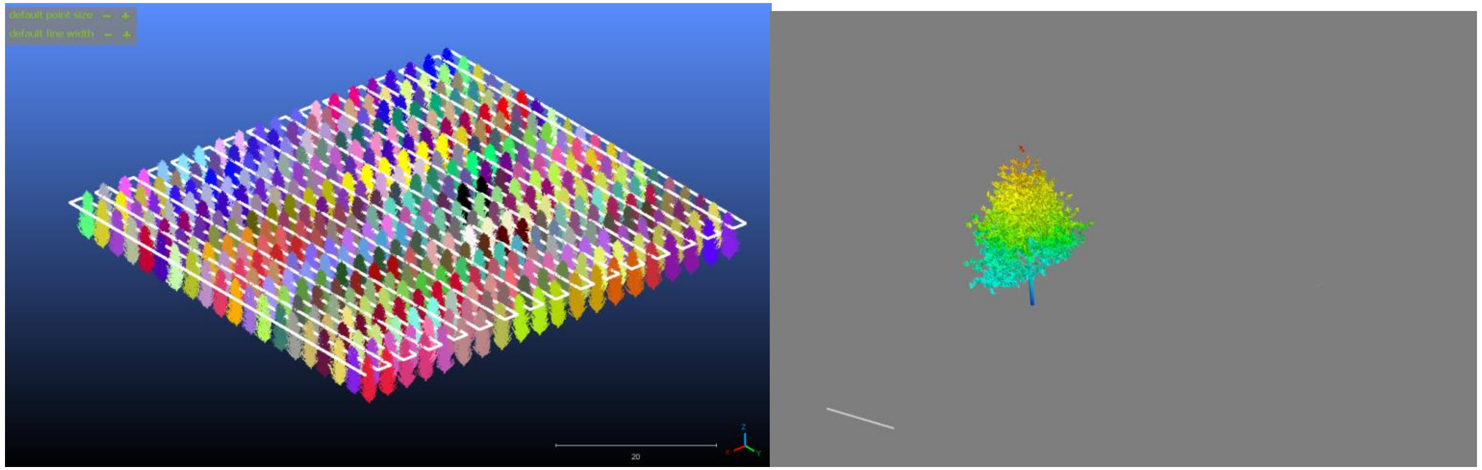


Two approaches for synthetic datasets creation

- High fidelity
- Naïve approach: copy-paste
- Computationally intensive



- Platform, scanner and trajectory settings dynamically controlled through HELIOS++ Python bindings
- Allows very detailed or very realistic point clouds
- Scaling up to stands (with needles) is extremely computationally demanding (2TB RAM)

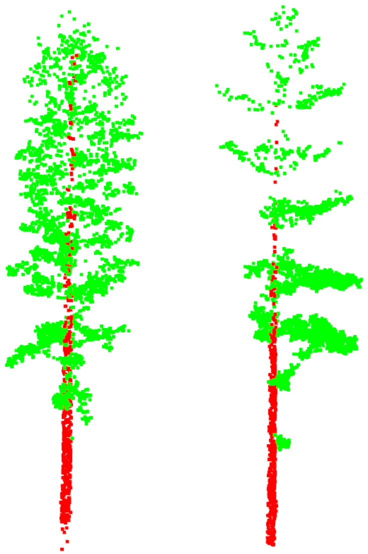


Ground Truth Trees

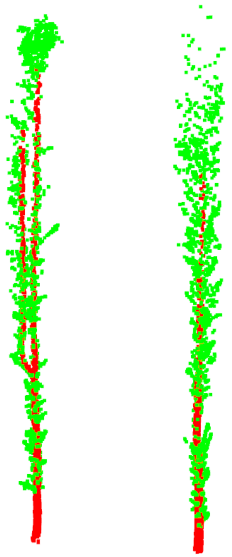
Airborne laser scanning (ALS)

Mobile ground-based scanning (MLS)

Mature Radiata Pine Trees
Tumut NSW, Australia



Pinus Caribaea spp.
Queensland, Australia



Recreational Forest,
Rotorua, New Zealand

Stem points

Foliage points

Learning-based Segmentation using Synthetic Helios Trees

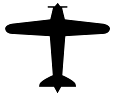
We use PointNet++ architecture[1] for individual tree segmentation

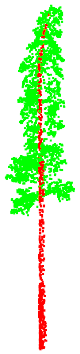
[1] Qi, C.R., Yi, L., Su, H. and Guibas, L.J., 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. NeurIPS, 30.

Quantitative Analysis - Results

Stem points

Foliage points



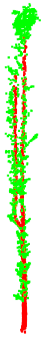


Segmentation performance on Radiata Pine trees from Tumut (Australia)

Model Training	Training on Real Dataset	Synthetic High-fidelity Training	Helios-backend Synthetic Training
Test IoU $_{foliage}$	**0.971**	0.838	**0.945**
Test IoU $_{stem}$	**0.528**	0.120	**0.247**

Stem points

Foliage points



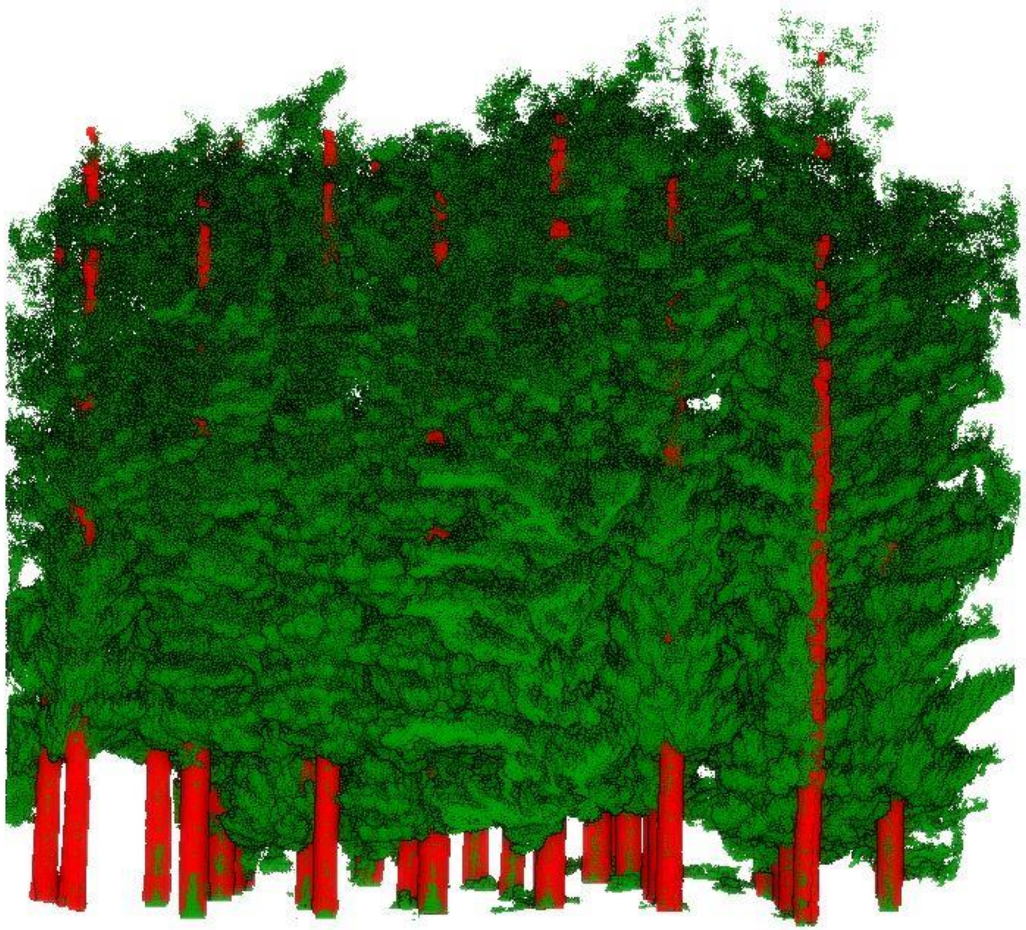
Segmentation performance on Radiata Pine trees from Scion Recreational Park (New Zealand)

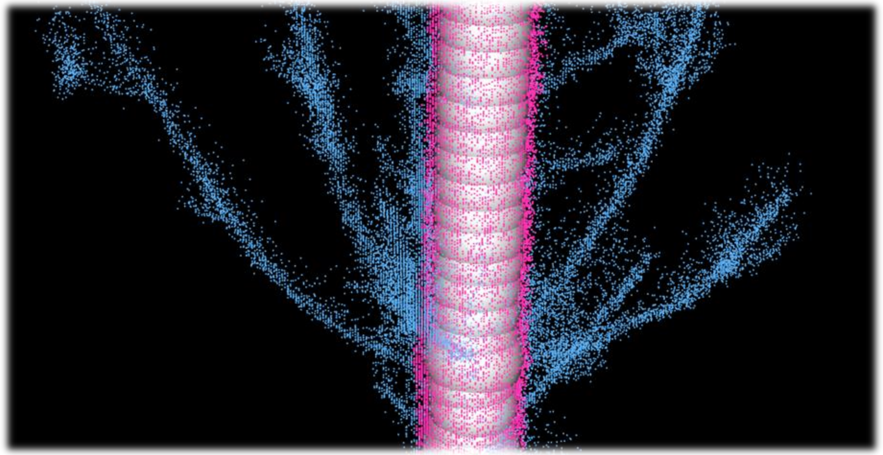
Model Training	Training on Real Dataset	Synthetic High-fidelity Training	Helion-backend Synthetic Training
Test IoU $_{foliage}$	**0.714**	**0.648**	0.623
Test IoU $_{stem}$	**0.633**	0.452	**0.572**

IoU = Intersection over Union (a measure of accuracy)

Results

- Quantitative Structural Modelling (QSM) results
- Strong results qualitative assessment as QSM is tuned to targets (DBH, volume etc.)





Transforming tree phenotyping | scion

Next-steps

- Unify benchmarking and datasets for 3D DL and QSM
- Exploring the potential Sim2Real data gaps
- Scale-up synthetic datasets
- Test alternative deep learning architectures
- Test hybrid QSM-3D DL approach

Acknowledgements

- We would like to thank
  - MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT HĪKINA WHAKATUTUKI for funding this research
  - RPBC® Radiata Pine Breeding Co for collaboration on this programme and access to trials.
  - Damien Sellier from Scion for his assistance with developing the synthetic tree pipeline
  - Peter Massam, Honey Jane Estarija, Warren Yorston and David Cajes from Scion for field data captures.
  - Timberlands Ltd., Manulife Forest Management (NZ) Ltd and New Zealand Forest Managers for access to their forests.
  - Interpine for additional data captures.

Forest SAT

Presented by Scion

- 2024 -

9-13 SEPTEMBER 2024

ROTORUA, NEW ZEALAND

forestsat.com/2024

Nicolò Camarretta

nicolo.camarretta@scionresearch.com

13 November 2023

www.scionresearch.com