# Primary Rainforest Mapping in Victoria 2018 - extent and type

Matt White, Khorloo Batpurev, Owen Salkin and Graeme Newell

May 2019



Arthur Rylah Institute for Environmental Research **Technical Report Series No. 309** 





#### Acknowledgment

We acknowledge and respect Victorian Traditional Owners as the original custodians of Victoria's land and waters, their unique ability to care for Country and deep spiritual connection to it. We honour Elders past and present whose knowledge and wisdom has ensured the continuation of culture and traditional practices.

We are committed to genuinely partner, and meaningfully engage, with Victoria's Traditional Owners and Aboriginal communities to support the protection of Country, the maintenance of spiritual and cultural practices and their broader aspirations in the 21st century and beyond.



Arthur Rylah Institute for Environmental Research Department of Environment, Land, Water and Planning PO Box 137 Heidelberg, Victoria 3084 Phone (03) 9450 8600 Website: <u>www.ari.vic.gov.au</u>

#### Citation

White, M., Bhatpurev, K., <sup>1</sup>Salkin, O. and Newell G. (2019). Primary Rainforest Mapping in Victoria 2018 - extent and type. Arthur Rylah Institute Technical Report Series No. 309. Arthur Rylah Institute for Environmental Research, Department of Environment, Land, Water and Planning, Heidelberg, Victoria.
 Natural Systems Analytics

#### Front cover photo

Warm Temperate Rainforest Goolengook Forest Block, Errinundra National Park, East Gippsland (Matt White).

© The State of Victoria Department of Environment, Land, Water and Planning 2019

(cc)

This work is licensed under a Creative Commons Attribution 3.0 Australia licence. You are free to re-use the work under that licence, on the condition that you credit the State of Victoria as author. The licence does not apply to any images, photographs or branding, including the Victorian Coat of Arms, the Victorian Government logo, the Department of Environment, Land, Water and Planning logo and the Arthur Rylah Institute logo. To view a copy of this licence, visit http://creativecommons.org/licenses/by/3.0/au/deed.en

Printed by Melbourne Polytechnic, Preston

ISSN 1835-3827 (print) ISSN 1835-3835 (pdf)) ISBN 978-1-76077-919-1 (Print) ISBN 978-1-76077-920-7 (pdf/online/MS word)

#### Disclaimer

This publication may be of assistance to you but the State of Victoria and its employees do not guarantee that the publication is without flaw of any kind or is wholly appropriate for your particular purposes and therefore disclaims all liability for any error, loss or other consequence which may arise from you relying on any information in this publication.

### Accessibility

If you would like to receive this publication in an alternative format, please telephone the DELWP Customer Service Centre on 136 186, email customer.service@delwp.vic.gov.au or contact us via the National Relay Service on 133 677 or www.relayservice.com.au. This document is also available on the internet at www.delwp.vic.gov.au

# Contents

Ackn	owledg	ements	iii		
Sum	Summary				
1	Introduction				
2	Methods				
2.1	Rainforest definitions				
2.2	The study areas				
2.3	Rainfor	Rainforest modelling			
	2.3.1	Dependent Data	4		
	2.3.2	Independent data	4		
	2.3.3	Data extraction and model development	4		
	2.3.4	Model validation	6		
	2.3.5	Model application, thresholding and conversion to polygons	6		
	2.3.6	Simplification, rendering and editing	6		
	2.3.7	Assigning rainforest type	6		
3	Result	6	9		
4	Discussion				
5	Conclusions and limitations				
6	References 1				

# **Acknowledgements**

The authors thank Tim O'Brien, Rod Taylor (Arthur Rylah Institute, DELWP), and Tom Hill and Natasha Reid (Forests, Fire & Regions DELWP) for their support, advice and assistance.

The authors also thank Arn Tolsma & Cindy Hauser (Arthur Rylah Institute DELWP) for making numerous improvements to an early draft of this report.

This project was funded by the Forest, Fire and Regions Group, DELWP as part of the Regional Forest Agreement forest values assessment program.

# Summary

#### Context

Rainforest arises in environmentally stable landscapes where high intensity fires have been absent for extended periods. Rainforest and related vegetation forms in Victoria are both naturally rare and susceptible to ongoing disturbances including bushfires, timber harvesting and disease.

Rainforest in Victoria is listed under the Flora and Fauna Guarantee Act 1988 and is protected from the direct impacts of timber harvesting activities.

This report has been prepared in support of analyses associated with the renewal of Victoria's Regional Forest Agreements.

#### Aims

Produce a data product to represent the current (circa April 2018) extent of primary rainforest across Victoria.

#### Methods

Machine learning algorithms were used to model the distribution of primary rainforest, using training data from known extant rainforest stands and multi-spectral satellite imagery.

The model formed the basis of a subsequent editing. Locations identified as potential rainforest by the model were checked, edited and augmented through the interpretation of high-resolution aerial photography.

#### Results

The model of rainforest extent reliably predicted the presence of rainforest stands in a variety of environmental and geographic contexts, however it tended to misclassify some areas of non-eucalypt vegetation, including mesophyll shrublands, urban gardens, *Acacia* stands (particularly *Acacia silvestris*), tree-fern dominated gullies and myrtaceous shrublands associated with streams and swamps. For this reason, a subsequent step of hand-editing was undertaken. This final editing process has been completed for the entire state apart from the Strzelecki Ranges which awaits further work.

We confirm that forest dominated by rainforest species is extremely rare in Victoria and comprises no more than 0.153% of the land surface of Victoria.

#### **Conclusions and limitations**

Where Victorian rainforest species comprise the canopy of the tallest vegetation stratum, rainforest extent has been accurately mapped for all regions of the state that support rainforest apart from the Strzelecki Ranges. The extent of rainforest in the Strzelecki Ranges is currently a modelled product that has not been further edited and reviewed, this is therefore indicative only.

The mapping of rainforest that exists or extends beneath an emergent eucalypt canopy (i.e. 'Mixed forest'), awaits further analysis of 'active' sensing data such as airborne or satellite-based laser scanning.

#### Published rainforest layer

The published rainforest layer can be obtained from the Victorian Spatial Datamart: RAINFOR "Rainforest Mapping for state-wide Victoria" Anzlic Id: ANZVI0803008605. 16 November 2019.

# 1 Introduction

Rainforest in Victoria is naturally rare as it can only arise in environmentally stable settings where high intensity fire has been absent for extended periods. Extensive clearing of rainforest for agriculture occurred in the 19<sup>th</sup> and 20<sup>th</sup> centuries and remaining stands are susceptible to the direct and indirect effects of disturbances such as fire, timber harvesting and disease (Ashton 1981; Busby and Brown 1994; Cameron 1992). As such, the extent of rainforest has been declining and warrants regular review.

Several types of rainforest community are currently recognised within Victoria. Four of these types are formally recognised in the state's Ecological Vegetation Class (EVC) typology: Cool Temperate Rainforest, Warm Temperate Rainforest, Gallery Rainforest and Dry Rainforest (Peel 1999). Additionally, Littoral Rainforest is also described and recognised as an 'endangered ecological community' by the Commonwealth EPBC Act. This latter vegetation type incorporates the EVC, Coastal Vine-rich Forest.

We combined modelling, using Sentinel and Landsat satellite image composites, species distribution modelling and interpretation of high-resolution aerial photography to create a map of rainforest extent and type across most of Victoria. Within the Strzelecki Ranges, rainforest extent and type have been modelled, but these outputs are yet to be confirmed and checked against high resolution aerial photography.

# 2 Methods

#### 2.1 Rainforest definitions

For the purposes of forest management in Victoria, rainforest is defined by the following criteria (see Vicforests 2015):

- 1. 'closed' (>70% projected foliage cover) broad-leaved forest vegetation with a continuous rainforest tree<sup>1</sup> canopy of variable height and with a characteristic diversity of species and life forms;
- 2. Where the broadleaved tree canopy of non-rainforest canopy species exceeds 10% projected foliage cover the forest is not considered rainforest.

This definition of rainforest covers all mature or primary rainforest. A broader ecological definition would also include forests where the broadleaved tree canopy of non-rainforest canopy species exceeds 10% projected foliage cover, provided that the first criterion has been satisfied. This expanded definition includes all vegetation currently referrable in Victoria to 'mixed-forest', 'seral rainforest' or 'secondary rainforest'. These are forests where rainforest exists beneath a eucalypt overstorey. Cool Temperate Mixed-forests are found in montane regions of the state<sup>2</sup>, have been described by Peel (1993) and are listed under the *Flora and Fauna Guarantee Act 1999* as a threatened ecological community. Other mixed-forest types are likely to be found in warm-temperate contexts, although their extent is likely to be restricted. Mixed forest has not been mapped in Victoria, largely because it is difficult to reliably identify these rainforest types using aerial photography as they are obscured by the eucalypt canopy. Consequently, their distribution has not been established with useful accuracy.

As this mapping project is informed by satellite imagery and aerial photography, it has been limited to mapping primary rainforest.

<sup>&</sup>lt;sup>1</sup> Rainforest canopy species are defined as shade tolerant tree species which can establish below an undisturbed canopy, or in small canopy gaps resulting from locally recurring minor disturbances, such as isolated wind throw or lightning strike, which are part of the rainforest ecosystem. Such species are not dependent on fire for their regeneration.

<sup>&</sup>lt;sup>2</sup> Notable the Central Highlands region and Errinundra Plateau

#### 2.2 The study areas

Two study areas encompassing the extent of rainforest in Victoria were delineated: Eastern Victoria and the geographically disjunct Otway Ranges (see Figure 1). This division was undertaken to facilitate the efficient 'handling' of the independent satellite data, which is data rich and finely resolved (to 10 m in several bands). Separate models were created for both study regions.

Overall the method progressed in two steps. First a spatially explicit and continuous model of the extent of rainforest was created, which was then thresholded and vectorized to make a first approximation of rainforest extent. Then a subsequent step of removing and modifying the first approximation polygons was undertaken. This process was informed by draping the initial predictions over high resolution aerial photography.



Figure 1. The location of the two study areas: 'Otway Ranges' in the west and 'Eastern Victoria' in the east. Areas shaded in green approximate the extent of rainforest in Victoria. The Strzelecki Ranges region shown as a blue outline is that region of the project where rainforest extent is modelled and has not been checked and edited and is considered less reliable.

#### 2.3 Rainforest modelling

Statistical modelling techniques such as regression establish mathematical relationships between dependent and independent datasets. These relationships can then be used to make predictions of the dependent variable in 'regions' beyond our existing knowledge, provided we have useful and more extensive independent data. In this study the presence or otherwise of rainforest is the dependent variable. It has only two possible states, 'presence' and 'absence' denoted by 1 or 0, however in the modelling analysis we treat this as a continuous variable in the regression analysis.

The independent variables for this model are remotely sensed variables derived from the Landsat or Sentinel satellite platforms. Multi-temporal, multi-spectral remotely sensed data derived from these platforms are particularly useful for modelling the extent of rainforest as these data:

- are spatially explicit and extensive
- are acquired regularly and as such will 'capture' changes consequent of logging and bushfire
- are rich in spectral data relevant to vegetation
- resolve at scales useful for planning purposes
- are objective and free from human bias and error.

#### 2.3.1 Dependent Data

We used exemplars from all five rainforest communities (i.e. Cool Temperate Rainforest, Warm Temperate Rainforest, Gallery Rainforest, Dry Rainforest, and Littoral Rainforest) for the eastern part of Victoria. Within the Otway Ranges, the exemplars were uniformly Cool Temperate Rainforest type as it is the only rainforest form present in this region.

The training data for rainforests were obtained using aerial photographic interpretation (API) with locations selected using contemporary imagery, and further informed by previous mapping. While every effort was made to minimise locational error, a small proportion of the 'rainforest presence' training data used in the Eastern Victoria model was later determined in the editing phase of the project to be incorrect as:

- 1. some stands identified sampled from 2013 aerial photography were subsequently destroyed by the East Gippsland Deddick River bushfire;
- 2. some stands that have been mapped as rainforest in the past were upon close inspection merely associated with or related to rainforest (e.g. Coastal Forest and Swamp Scrubs in Croajingalong National Park).

These 'false training positives' represented less than 5% of the positive training set for East Gippsland.

Rainforest absence data were derived from both the allocation of random or background absence (*sensu* Liu et al. 2016), and the deliberate placement of a significant number (approximately 10,000 in Eastern Victoria, and 3000 in the Otway Ranges) of absences in a range of forest types including plantations and regrowth and many within 50m of the rainforest training polygons. The latter were included to accurately render the rainforest, sclerophyll forest boundary. In all, approximately 247,000 training sites were supplied to the eastern Victorian rainforest model comprising approximately 55,000 rainforest sites. The Otway Ranges model was informed by approximately 17,000 training sites, 2,300 of which were rainforest sites. These training data became 'targets' in the separate machine learning models.

#### 2.3.2 Independent data

We used Sentinel 2 and Landsat thematic mapper products as independent data for the model. Both Sentinel and Landsat are regularly acquired and free to access. The Sentinel satellite data were processed and accessed using the Google Earth Engine (GEE) platform. A mosaic for each of the study areas was created from the median of each of 10 spectral bands for each pixel, obtained from the cloud free set of images over the summer 2017-2018.

Each mosaic comprised bands 2, 3, 4, 5, 6, 7, 8, 8a, 11, 12. Bands 5, 6, 7, 8a, 11, 12 each have a native resolution of 20m, and these data were resampled to 10 m using bilinear interpolation to coincide with resolution of Bands 2, 3, 4 and 8. Band ratio values for each pixel were then calculated from these Sentinel data as independent inputs to the modelling to subdue the effect of disparate illumination due to terrain and satellite position (see Table 1).

In addition to the Sentinel data, composite products from the 32-year long Landsat - 5 band (Thematic Mapper) satellite dataset were used to 'inform' the models of forest dynamics and disturbance. One variable was created from the 32-year cloud free set of summer images from the 75<sup>th</sup> and 25<sup>th</sup> percentiles of the red and green bands (bands 3 and 2 respectively). These were incorporated into a single variable using a normalised difference ratio approach (see Table 1). The final list of independent data is provided in Table 1.

#### 2.3.3 Data extraction and model development

The locations of the dependent data (both presence and absence data) were used to extract the coincident values of the independent satellite ratio data to create the modelling dataset. We used the CLUS system (Struyf *et al.* 2011) to develop a regression tree model that employed bagging and ensembling for each of our two study areas. The goal of regression trees is to predict the target (or dependent variable) by the recursive partitioning of the input (independent) variables. Hence, each leaf in the tree represents a value of the target variable given the values of the input variables and can be represented by the path from the root to the leaf (Friedman 2001). Specifically, we used predictive clustering trees (*sensu* Kocev *et al.* 2007), which is a type of regression tree that generalizes learning trees as cluster hierarchies. Each decision node within the tree is supplied with the full set of the independent variables from which a partitioning test is applied. Bootstrap aggregating (or 'bagging'), which is like model averaging (Breiman 1996), was used to improve the accuracy of predictions. In all, 50 bootstrap sub-samples each comprising 33% of the entire modelling dataset were developed. The resultant suite of models were averaged to produce a consensus or ensemble model through model voting.

#### Table 1. List of independent (predictor) satellite variables.

Location of images and training data Aire Valley, Otway Ranges Green markers are Rainforest presence data Blue markers are Rainforest absence data	
Derived satellite variables	
The ratio of Band 2 (green) with Band 3 (Red)	
The ratio of Band 2 (green) with Band 6 (Near Infrared 1)	
The ratio of Band 2 (green) with Band 7 (Near Infrared 2)	
The ratio of Band 2 (green) with Band 8 (Shortwave Infrared band 1)	
The ratio of Band 2 (green) with Band 8a (Shortwave Infrared band 2)	
Band 5 (Red Edge) pansharpened with Band 3 (Red)	
((75th %tile B2 / 75th %tile B3)–(25th %tile B2 / 25th %tile B3)) * ((75th %tile B2 / 75th %tile B3)+(25th %tile B2 / 25th %tile B3))	

\* Landsat TM, 32 year mean data. All other imagery from Sentinel

#### 2.3.4 Model validation

Two independent ensemble models were built for each study area; one that 'bootstrap sampled' from all the available presence and absence data, and another that was restricted to sub-sampling from 90% of the presence and 90% of the absence data. The latter model was tested in terms of its capacity to predict the remaining 10% of the data that was withheld from the modelling process. This test indicated the underlying performance of the final model and the degree to which it can be generalised.

#### 2.3.5 Model application, thresholding and conversion to polygons

The modelled relationships between the independent and dependent data formulated by the consensus or ensemble model were used to create continuous pixel-based maps of the probability of the presence of rainforest at 10m resolution across both study areas (see Figure 2 - upper right panel). A suitable threshold from the continuous data to delimit rainforest was determined in both study areas by draping the models across aerial photography. In both regions a probability range of 0.49-1 was selected to ensure that false negatives were minimised. Then the raster data that lay within this continuous range of values was then converted to polygon format, or 'vectorized'.

#### 2.3.6 Simplification, rendering and editing

Following the hand editing the polygons were subjected to post-geoprocessing which included an outline smoothing process and processing to remove holes within the mapped rainforest extent if they were less than 400m<sup>2</sup>. All polygons beyond the Strzelecki Ranges were evaluated against the most recent, high resolution aerial photography. Polygons were variously reshaped or removed in this process. Figure 2 shows a sample of the development process of the final polygon extent, from the raw modelled outputs to the edited polygon theme.

#### 2.3.7 Assigning rainforest type

The rainforest types (i.e. Cool Temperate, Warm Temperate, Dry and Gallery) were assigned to polygons by numerical analysis. Spatially explicit presence-absence species distribution models (see Liu *et al.* 2011) of the dominant rainforest tree species in Victoria were used to determine the 'most likely' rainforest type that would be present at any location. These models had previously been developed from vegetation plot data. Species models of each rainforest tree were aggregated by the type to which they have the most fidelity (see Table 2) in accordance with Peel (1999). For rainforest types with more than one 'characteristic species', the maximum probability of any of the member species was resolved at each pixel. The most likely type was then determined by maximum 'type probability' at each pixel (see Figure 3) to produce a spatially explicit map of the most likely type. Rainforest types were assigned by determining the 'majority type' intersecting each polygon. Some hand editing of polygons was undertaken to split polygons that straddled the boundaries between Cool-temperate Rainforest and Warm-temperate Rainforest in East Gippsland. Littoral rainforest polygons were determined based on their proximity to estuaries and open coastline.

Rainforest Type	Dominant rainforest genera / species		
Cool Temperate	Nothofagus cunninghamii		
	Atherosperma moschatum		
	Elaeocarpus holopetalus		
Warm temperate	Syzygium smithii		
	Acronychia oblongifolia		
	Elaeocarpus reticulatis		
Dry	Brachychiton populneus		
Gallery	Tristaniopsis laurina		
Littoral	Various / idiosyncratic (stands labelled by coastal or near coastal context)		

Table 2.	Rainforest tree	species used t	to characterise	rainforest E	Ecological Ve	detation Classes
						90.000.000000



Figure 2. Box A shows a sample region of forest around the Melbourne Ladies College campus in far east Gippsland near the NSW border. The inset in this same panel shows the approximate location of the images. Box B shows the results of the continuous (probabilistic rainforest extent modelling). Box C shows the thresholded model, converted to polygons and clipped by the Victorian state border. Box D is the final model following a hand editing process and smoothing to remove small gaps internal to stands (Note this final step was not completed for the Strzelecki Ranges).



Figure 3. 'Most likely' rainforest type determined for forest in eastern Victoria. Panel A (above), shows what type of rainforest would be most likely if rainforest occurred there, i.e. even where the presence of rainforest is impossible or unlikely there is a 'most likely' rainforest type based on the dominant tree species. Panel B (below), displays the same model constrained by the rainforest extent model. This depicts the likely rainforest type in places where rainforest currently exists or may have existed in the recent past.

# 3 Results

The R<sup>2</sup> or, general 'fit' of the rainforest models for both the Otway Ranges and Eastern Victoria is extremely high at 0.9909 and 0.9905 respectively. Model fit remained high when tested against the validation hold-out; at 0.9668 and 0.9608 respectively. This degree of accuracy is not unusual for models of this nature where a geographically rare landcover type is contrasted with all other landcover types using covariates acquired from satellites (Papas *et al.* 2016). The key question of such modelled products is however whether it is 'fit-for-purpose'. This frequently relates to:

- How well the exemplars sample the covariate variation represented in each class (i.e. the classes 'rainforest' and 'not rainforest).
- How well the model operates in the restricted ambiguous 1% of the landscape where model performance may be critical, such as accurately identifying narrow strips of rainforest in poorly illuminated gullies.

In this project the modelling has essentially been used to highlight and identify places that may support rainforest and warrant further checking or 'truthing' using high resolution aerial photography. As such, the success or otherwise of the modelling is partially qualitative as it focusses an 'efficient searching' and verifying spatial precision, particularly within the context of an elevated sensitivity to false absences (i.e. missing extant rainforest stands). In general, the model over predicted the extent of rainforest by misclassifying significant areas of various vegetation types. In general, these were non-eucalypt dominated vegetation types such as mesophyll shrublands, urban gardens, exotic broadleaf trees in rural areas, *Acacia* stands (particularly those dominated by Blackthorn *Acacia silvestris*), tree-fern dominated gullies and myrtaceous shrublands associated with streams and swamps. These were removed by a hand editing process. However, significant regions of these non-rainforest types persist in the modelled rainforest extent in the Strzelecki Ranges.

Tables 3 and 4 show the 'confusion matrices' for the separate regional rainforest models. The tables show the apparent performance of the model to correctly classify the model training data in its raw raster form, as initial unedited, binary-vectorized form, and in its final form post editing. The decline in apparent accuracy for the final model is attributable to several factors including the model's spatial precision (i.e. note the decline in false negatives if we merely buffer the final model by 20 metres), the inevitable number of false negatives that are the consequence of the use of random background 'absence' data<sup>3</sup>, and the number of rainforest training points that were in error to begin with (many of which appear as 'false positives'; see section 2.2). We suggest that outside the Strzelecki Ranges region (see Figure 1) the editing process has eliminated almost all error.

The project has revealed that there is up to 34,670 hectares of rainforest remaining in Victoria, of which approximately 30,200 hectares has been confirmed through rigorous checking using high resolution aerial photography. An example of the results from the modelling and mapping process are shown in Figure 4, these are contrasted with the pre-existing EVC rainforest from earlier EVC mapping studies.

The project has also revealed several outlying and disjunct rainforest stands in regions previously considered to be outside the 'environmental envelope' suitable for the development of rainforest communities. For example, this includes patches of Cool-Temperate Rainforest in the headwaters of Livingstone Creek south of Omeo.

In addition, we have found that rainforest has significantly contracted over the last 20 years from known locations. There has been extensive loss of both primary and secondary rainforest as a consequence of bushfires, and significant stand losses and contractions have occurred on Wilson's Promontory (i.e. losses due to landslides and drought), the Blue Range (e.g. Marysville and Lake Mountain area - the only area supporting rainforest north of the Great Dividing Range), the catchment of the O'Shannassy River near East Warburton, and the Rodger River catchment in East Gippsland. It also seems likely that primary rainforest is no longer extant in the Mount Disappointment area.

<sup>&</sup>lt;sup>3</sup> If we assume Rainforest occupies something approaching 2% of our study areas, then the use of random background absence may include up to 4000 erroneous absences in Eastern Victoria and up to 300 erroneous absences in the Otway Ranges.

# Table 3. Error rates on training data for the Eastern Victorian Rainforest extent model (confusion matrices)

	Raster Model (threshold at 0.49)	Unedited polygons derived from model	Final edited Rainforest polygons	Edited polygons Rainforest polygons buffered by 20m
True Negatives (not-rainforest exemplars)	189525	190116	198254	195430
True Positives (rainforest exemplars)	53262	32533	29148	32364
False Negatives (not-rainforest exemplars)	1634	4616	8001	4785
False Positives (rainforest exemplars)	2081	19237	11099	13923
Sensitivity	0.9702	0.8757	0.7846	0.8711
Specificity	0.9891	0.9081	0.9469	0.9334

# Table 4. Error rates on training data for the Otway Ranges Rainforest extent model (confusion matrices)

	Raster Model (threshold at 0.49)	Unedited polygons derived from model	Final edited Rainforest polygons	Edited polygons Rainforest polygons buffered by 20m
True Negatives (not-rainforest exemplars)	14357	13768	13825	13459
True Positives (rainforest exemplars)	2329	2154	1813	1918
False Negatives (not-rainforest exemplars)	34	209	550	445
False Positives (rainforest exemplars)	28	617	560	926
Sensitivity	0.9856	0.9115	0.7672	0.8116
Specificity	0.9980	0.9571	0.9610	0.9356



Figure 4. Contrasting images of EVC Rainforest mapping. The left panel shows the pre-existing rainforest extent, without type. The right panel displays rainforest mapping produced from this project. The location is the Coast Range, East Gippsland, as displayed by the black rectangle on the map of Victoria. This region to the east of Goongerah is centred on the 'Frosty Hollow' rainforest, the largest contiguous stand of primary rainforest in Victoria.

# 4 **Discussion**

Rainforests as discrete plant communities can only develop in environmental settings that are stable and moist over very long periods of time (i.e. generally well in excess of 100 years), and that are free from physical disturbances such as clearing and fire events (Ashton 1981; Busby and Brown 1994; Cameron 1992). Given the time required to develop rainforest communities, and their sensitivity to various forms of disturbances, rainforests are afforded a high status in terms of conservation and reservation.

The composition (i.e. species mix) of rainforest communities will naturally differ across different climatic regimes, edaphic and topographic settings. It is therefore necessary to clearly discriminate and demarcate the distribution of these communities, to minimise the risks of inadvertent disturbances.

While the machine learning models developed here successfully detected rainforest vegetation across eastern Victoria and Otway Ranges, the models tended to over-predict the extent of rainforest into other, typically, non-eucalypt dominated vegetation. This is not an entirely surprising result, given that the models are essentially a mathematical competition between the presence and absence of a feature or target. Therefore, for modelling algorithms to operate with the high levels of discrimination required to distinguish between subtle changes in canopy composition, they need to be provided with an extensive range of examples of what is 'not' the target (i.e. the naturally wide variance in the non-rainforest 'absence' data). In the context of this project, it was difficult to provide a complete and representative set of exemplars or training data for all non-rainforest types across each study area.

Given the timeframes for this project and the requirement for a reliable data product, it was considered that the best way to proceed was to threshold the models to levels that were unlikely to omit rainforest systems from mapped products, and then to remove polygons that were clearly not rainforest, and complete hand editing on mapped rainforest where this was required. This was accomplished by draping the modelled output (i.e. thresholded and in vector form), over high resolution contemporary aerial imagery. Although this approach was more time-consuming than originally anticipated, it did provide a high level of quality assurance. However, one unintended consequence of the time-consuming of hand editing was that verified mapped products could not be produced for the Strzelecki Ranges during the timeframe of the study.

Comparisons between the current rainforest extent (as incorporated in the current native vegetation extent dataset), and the new rainforest model clearly indicate that in general, there is a high level of accordance between the older mapped and newer modelled rainforest datasets. However, there are regions where the new models suggest that the existence of rainforest may have been below the detection limit / mapping resolution in the 1990's. In these instances, the 2019 product has detected regions of rainforest vegetation that were not mapped previously. This is also coupled with a considerable improvement in the detail of the new line work delineating rainforest stands, which is unsurprising given the improvements in both aerial photography and satellite imagery over the last two decades. Given this tighter delineation of rainforest areas, buffering of these data may be useful to produce data products suitable for regulatory purposes, while some simplification or other geo-processing operations may be useful prior to the inclusion of these data in native vegetation datasets.

# 5 Conclusions and limitations

Where Victorian rainforest species comprise the canopy of the tallest vegetation stratum, rainforest extent has been accurately mapped for all regions of the state that support rainforest communities, apart from the Strzelecki Ranges.

The extent of rainforest in the Strzelecki Ranges is currently a modelled product that has not been further edited and reviewed and is therefore indicative only.

The mapping of rainforest that exists or extends beneath an emergent eucalypt canopy (i.e. 'Mixed forest'), awaits further analysis of 'active' sensing data such as airborne or satellite-based laser scanning.

## 6 References

- Ashton D. H. (1981). Fire in tall open-forests (wet sclerophyll forests). In: Fire and the Australian Biota (eds A. M. Gill, R. H. Groves and I. R. Noble) pp. 339-66. Australian Academy of Science, Canberra.Breiman, L. (1996). Bagging predictors. Machine Learning 24: 123–140.
- Busby J. R. & Brown M. J. (1994). Southern rainforests. In: Australian Vegetation (ed R. H. Groves). University Press, Cambridge.
- Cameron D. (1992). A portrait of Victoria's rainforests: distribution, diversity and definition. In: Victoria's Rainforests: Perspectives on Definition, Classification and Management (eds P. Gell and D. Mercer) pp. 13-96. Monash Publications in Geography No. 41. Monash University, Melbourne.
- Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. Annals of Statistics 29 (5): 1189–1232.
- Kocev, D., Vens, C., Struyf, J., and Dzeroski, S. (2007). Ensembles of multi-objective decision trees. In: Kok J.N., Koronacki J., Mantaras R.L., Matwin S., Mladenič D., Skowron A. (eds) Machine Learning: ECML 2007. ECML 2007. Lecture Notes in Computer Science, vol 4701. Springer, Berlin, Heidelberg
- Liu, C., Newell, G., & White, M. (2018). The effect of sample size on the accuracy of species distribution models: Considering both presences and pseudo-absences or background sites. Ecography 42 (3): 535-548.
- Liu, C., White, M. & Newell, G. (2011). Measuring and comparing the accuracy of species distribution models with presence–absence data. Ecography 34, 232–243.
- Papas, P., White, M., Cant, B., Griffioen, P., Crowther, D. and Cook, D. (2016). Predicting the occurrence of seasonal herbaceous wetlands in south-east Australia. Arthur Rylah Institute for Environmental Research. Technical Report Series No.271. Department of Environment, Land, Water and Planning, Heidelberg, Victoria.
- Peel, B. (1999). Rainforests and cool temperate mixed forests of Victoria. Department of Natural Resources and Environment, East Melbourne, Victoria 3002 Australia
- Struyf, J., Zenko, B., Blockeel, H., Vens, C., Džeroski, S. (2011). CLUS: User's Manual. https://dtai.cs.kuleuven.be/clus/publications.html.
- VicForests (2105). Rainforest Identification Instruction, Version 2. http://www.vicforests.com.au/static/uploads/files/vicforests-rainforest-identification-instruction-v2-0wfhuonwjqqu.pdf

www.delwp.vic.gov.au

www.ari.vic.gov.au