Multi-temporal Land Cover and Native Vegetation Extent for Victoria

M. White, P. Griffioen, and G. Newell

March 2020

Arthur Rylah Institute for Environmental Research
Technical Report Series No. 311

Arthur Rylah Institute for environmental research

Arthur Rylah Institute for environmental research 1970 - 2020 50 YEARS of science that matters

VICTORIA State Government | Environment, Land, Water and Planning

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: www.ari.vic.gov.au

Citation: White, M., Griffioen, P. and Newell, G. (2020). Multi-temporal Native Vegetation Extent for Victoria. Arthur Rylah Institute for Environmental Research Technical Report No 311. Department of Environment, Land, Water and Planning.

Front cover photo: Steve Sinclair, ARI.

Printed by Melbourne Polytechnic, Preston

Edited by Tim O'Brien

ISSN 1835-3827 (print)
ISSN 1835-3835 (pdf))
ISBN 978-1-76105-018-3 (print)
ISBN 978-1-76105-019-0 (pdf)

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

Multi-temporal Land Cover and Native Vegetation Extent for Victoria

Matt White, Peter Griffioen, and Graeme Newell

Arthur Rylah Institute for Environmental Research
123 Brown Street, Heidelberg, Victoria 3084

Arthur Rylah Institute for Environmental Research
Technical Report Series No. 311

Arthur Rylah Institute for Environmental Research
Department of Environment, Land, Water and Planning
Heidelberg, Victoria

Acknowledgements

The authors acknowledge the support of a 2015 research grant from the South Australian Government which was used to develop and refine methods appropriate for landcover and native vegetation extent modelling across the State of South Australia.

Specifically, the authors thank Matthew Miles and Matthew Royal of the South Australian Department of Environment and Water (DEW) and Brad Page (formerly DEW) South Australian Department of Primary Industries. Additionally, the authors acknowledge the data science support provided by Geoscience Australia, particularly Alexis McIntyre (now Department of Agriculture, Water and the Environment) and Norman Mueller.

The authors also thank the Biodiversity Division staff within the Victorian Department Environment, Land, Water & Planning particularly Dave Adams, David Attard and Anne Buchan, for support to pursue the development of equivalent data products for the State of Victoria which is the primary focus of this report.

Contents

Tables

Figures

Summary

Context:

Consistent information on the extent and spatial arrangement of native vegetation is essential not only for DELWP, but also for many other management agencies, NGOs and many interested citizens and community groups. The production of such datasets has previously been a laborious process involving on-ground ecological field work and botanical analyses, followed by aerial photographic interpretation by skilled personnel. Projects of this nature were slow to conduct, expensive, and often lacked consistency and transparency of the processes undertaken. Mapping in this way was not easily repeatable or updatable without a reinvestment of a similar magnitude. The need to provide timely, consistent and cost-effective reporting on the dynamic changes on the spatial arrangement of native vegetation is the key driver for this project.

Aims:

1. Improve and streamline the processes for reporting on the dynamic changes in the exent of native vegetation and its spatial arrangement and configuration.
2. Deliver products that are more relevant for conservation planning and reporting, along with other needs of the NRM sector.

With the above focus, it is important to recognise that the admixture of surrounding land cover types and land uses is highly relevant to conservation objectives. For example, adjacent land cover types to native vegetation can be either benign / supportive, or alternatively neutral or even antagonistic to long-term conservation. Access to this level of contextual information has been a major knowledge gap for effective landscape scale planning and management for many decades, both at broad / state-wide and local scales.

Methods and Scope:

This significant project covered the spatial extent of both Victoria and South Australia, and was partially funded by DELWP, and from a research grant from the SA Government. The approach taken is multi-disciplinary, and integrates remote sensing, ‘big data’ analytics / machine learning (ML) algorithms, in addition to ecological data and knowledge, and historical patterns of land use. These of fields of study are all advancing rapidly with improved technologies, data handling, computational processes and improved algorithms, which will undoubtedly provide future opportunities for refinements, and the inclusion of additional / ancillary data.

Results:

This report provides details on the development and provision of a suite of new data products / resources for DELWP: the Victorian Land Cover Layers 1985–2019. These products are a series of six spatial datasets that can either be viewed independently or sequentially over time from the late 1980’s until late 2010’s. The data outputs have been developed around a logic and method that is transparent, updateable and repeatable. This document details the methods used to develop these products, along with limitations and caveats that apply to the data products.

Conclusions and implications:

These landcover datasets may be useful for conservation planning needs, and for NRM practitioners, but may not be suitable for other planning and resource management applications. These products from machine-learning models of remote-sensed data are complex and nuanced, and may differ substantially from maps formerly produced using conventional cartographic techniques. The resolution of the remote sensed data means that the model may struggle to adequately reflect the extent of scattered trees, or those with sparse foliage (e.g. *Allocasuarina* spp.). The processes used required summarizing remote sensed data over multiple years into epochs to report on changes between-epochs. Consequently, the outputs cannot detect or report on dynamic or transient land cover changes within-epochs (i.e. ~5 year time period).

1 Introduction

Datasets providing information on native vegetation extent provide critical information for DELWP and many other management agencies, as well as NGOs, many community groups and interested citizens. These data are essential to appreciating the current spatial distribution and arrangement of native vegetation, but also in setting clear priorities for management actions and expenditure.

Importantly, the extent of native vegetation is dynamic, and it will be critical for DELWP to develop the capacity to capture information of this nature, and to track changes in vegetation over time. Native vegetation extent mapping has been undertaken infrequently over the last two decades or more, and has generally been coincident and opportunistic, and often aligned to specific policy development needs (e.g. Regional Forest Agreements, changes to Native Vegetation Regulations, etc.). Generally, these mapping programs have captured and represented native vegetation cover (under several different schema), but they generally have not clearly identified or indicated the surrounding non-native vegetation cover types. Understanding the spatial context, as well as the extent and arrangement of land cover types is critical for strategic planning for conservation management. Land cover types (and associated land uses), can be either benign / supportive of conservation objectives, or alternatively neutral or even antagonistic to these goals. Developing data products that provides this contextual information has been a major knowledge gap and impediment to effective conservation planning, both at broad / state-wide and local scales.

The original large-scale project was initiated through a research grant to Arthur Rylah Institute to develop and refine appropriate methods for the South Australia government, building upon some preliminary work in Victoria for developing multi-temporal land cover time series. This current project funded by DELWP built upon the previous project and sought to refine the data products to suit its specific immediate requirements of the Biodiversity Division.

The information provided in this report relates to the dependent and independent data, the modelling methods, and the limitations and caveats that apply to the data products for both South Australia and Victoria. The results presented however relate more specifically to the spatial subset of the multi-temporal land cover dataset provided to the Biodiversity Division, DELWP, that extends across Victoria.

2 Aims

The general aim of the project reported here was to develop and provide a suite of new data products / resource for DELWP: the Victorian Land Cover Layers 1985–2019. This is a series of seven spatial datasets that can either be viewed independently or sequentially over time from the late 1980’s until late 2010’s. These data outputs are consistent with the logic and methods used for the development of similar products for South Australia, but have intentionally been truncated in the number of output land cover classes to report of the primary land cover features across Victoria across a greater than a 30-year time frame.

This document seeks to document the underlying logic, and the technical processes that were devised to enable the production of these new datasets. Additionally, caveats and limitations of this process and products are also provided.

It is also important to recognize that this innovative project was undertaken by integrating data and modelling from several different scientific fields including remote sensing, ‘big data’ analytics / machine learning (ML) algorithms, as well as ecological data. All these fields are advancing rapidly with improved technologies, data handling, computational processes and improved algorithms. This documentation of the current logic / process will provide future opportunities for new approaches and additional / ancillary data relevant to this and related spatial modelling tasks.

3 Methods

A range of methods were employed in this project to maximize the consistency, accuracy and repeatability of the vegetation extent and land-cover models. Significant features of the modelling processes include:

- the inclusion of new earth observation data commissioned from Geosciences Australia (GA). These data were processed by GA using the National Computing Infrastructure to ARI / DELWP's specifications.
- the use of multi-class attributions to on-ground training locations of landcover types (i.e. exemplars). Using multi-target regression tree modelling processes allows for the inductive transfer of information from one land cover class to all other classes when all classes are modelled simultaneously within a single, multi-target model. This modelling framework provides significant advantages to modelling a single target or class independently (e.g. 'native tree'), and subsequently attempting to combine and reconcile the numerous models for all the other classes.
- the use of ensemble methods, along with stratified 'bagged' sampling, improves model accuracy, reduces sample biases, and explicitly allowed for a characterisation of within-model uncertainty.
- the development of a coherent non-temporal model that can be applied to any relevant temporal remote-sensed dataset that meets the current specifications for remote-sensed data.
- the development of software for vetting of exemplar data based upon the level of model variance or discrepancy in class assignation.
- the development of inter-epoch rules to consistently provide the 'most-likely' land cover classes.

Study area and study period description

The spatial extent of the study area was the geographic extent of the Australian States of South Australia and Victoria. The temporal extent of the study area is the period 1987 until 2019 summarised by seven 'overlapping' data epochs:

- 1985-1990
- 1990-1995
- 1995-2000
- 2000-2005
- 2005-2010
- 2010-2015
- 2015-2019

Epochs intentionally overlapped the next in the series by one year, to ensure a sufficient number of cloud free images would be obtained to calculate the relevant pixel statistics within each epoch, and for each season. In this way, each epoch represents a rolling average of reflectance values relevant to a broad time period, and not a specific time period (e.g. median year) within the epoch. Most epochs covered a five year period, apart from the most recent that spanned a four year interval.

Model inputs

Land-cover classification rationale

The intent of this project differed from previous land cover programs at state-wide and national scales that have focused upon land use / agricultural productivity / economic activity. The rationale for this project was to develop time series of native vegetation and land cover classes that provide more contextual and nuanced information relevant to conservation planning and natural resource management.

Previous land cover modelling projects have generally been produced for a single point in time, while the intent of this project was to develop a framework and to deliver products that were temporally consistent, and which enabled reporting on land cover changes over time, including the extent of native vegetation. For this and other reasons, careful consideration was given to developing data structures and systems that could be added to and maintained over time. This was to facilitate the ease for future updates to the sequence of landcover products, in addition to providing the considerable potential for adapting this framework for the development of ancillary and related data products in the future (i.e. multi-temporal and spatial modelling). Prior experience has demonstrated to the authors that while modelling of a single or direct 'target' can be a productive strategy for model development in some instances, a more adaptable and powerful approach is to use a 'multi-target' modelling approach which delivers more consistent, integrated and reliable data products.

This approach used within the project involved a hierarchy of forms of data and modelling steps. The primary data for modelling in this project were observations of land cover types (i.e. exemplars – blue box, Figure 1), which were made for all exemplar land cover types and across all epochs. These observations were used to develop a large 'multi-target' model for 40 different land cover attributes. Each of these interrelated / co-joined models provided a probabilistic likelihood for a particular landcover attribute being present at a pixel within each epoch (see orange box – Figure 1). Importantly, modelled outputs of land cover attributes will understandably 'overlap' and will therefore not be mutually exclusive. For example, a single pixel within a patch of native forest would be expected to have high probabilities for several different attributes such as native, land_vegetated, land_vegetated_woody, and treed_native attributes; and lower probabilities for the other 36 land cover attribute probabilities.

The next step in the production of the mapped land cover outputs requires the definition of the 'target' land cover classes. For the project reported here relevant to Victoria, 19 land cover classes were agreed upon between the Knowledge and Decision Systems Branch and ARI staff as the 'target' classes (Table 1). The combination of the probabilities for each land cover attribute at each pixel contributed to the ultimate classification of final target land cover type. Further details for each of the steps are provided later in this document.

Figure 1 General modelling logic and relationships between observations of land cover classes, the multiple attributes that can be assigned to the observation, and the target land cover class.

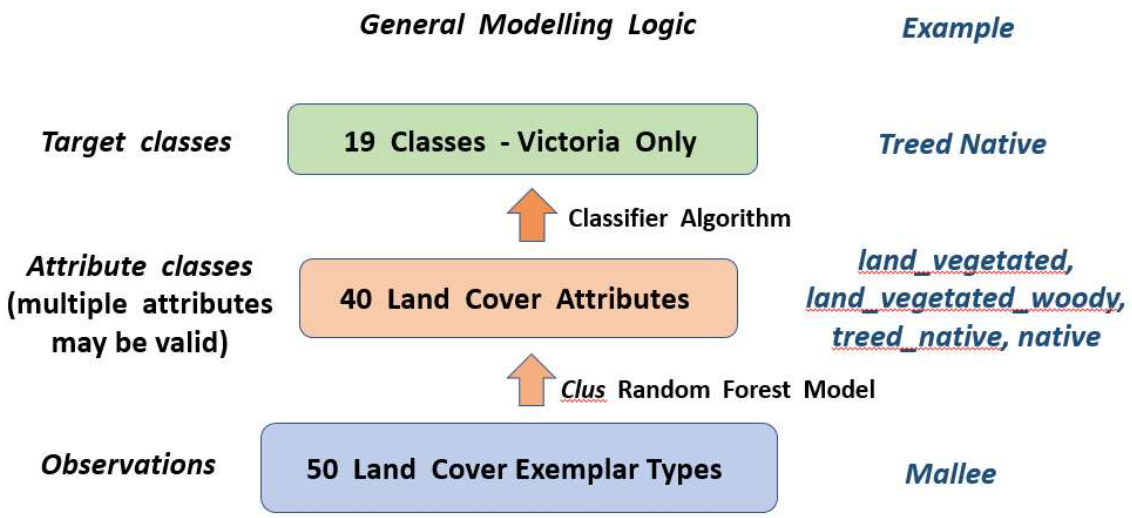


Table 1 Nineteen land cover classes defined as targets for Victorian land cover outputs.

These units represent aggregations of 40 land cover attributes briefly outlined in Table 2 and Figure 1, and are the ultimate product from ensemble modelling and followed by the application of the post-processing methods. Definitions for these classes are provided in Appendix 1.

Victorian Land Cover Classes	
Water	Wetland perennial (native)
Natural low cover (Bare-ground – Native)	Wetland seasonal (ephemeral – native)
Disturbed ground (Bare-ground - Not Native)	Treed Native vegetation
Built Environment	Native Shrub-land
Scattered native trees	Hardwood plantation
Urban area	Conifer plantation
Native pasture / Grassland	Other exotic tree cover
Exotic pasture / Grassland	Horticulture/irrigated pasture and crop
Dryland cropping	Saltmarsh vegetation
	Mangrove vegetation

Dependent data

Statistical and machine-learning (ML) modelling rely upon establishing relationships between two types of data: dependent and independent data. Dependent data (i.e. the response variable), are generally the observations that have been made. These generally become the focus or the 'target' of a model that is developed using the independent data (i.e. the predictor datasets).

The dependent data in this study were observations (i.e. precise locations with x & y coordinates), of representative for 40 land cover attribute classes (Table 2), that are compiled from observations of 50 types of land cover exemplars. Each exemplar observation was effectively 'time-stamped', in that the observation related to only one of the 6 epochs across the time series. In practice these observations may be static, and therefore an observation is relevant to all epochs across the series (e.g. always a 'treed_native' in a National Park). Alternatively, the observations at a pixel may change over the 30+-year time period; e.g. treed_native may be cleared to become dryland_cropping, which subsequently changes to built environment when new suburbs are established. All observation data were held within a separate lookup table within the Exemplar Database. These attributes are the features that are being modelled mathematically, and the classes are organized predominantly within a semi-hierarchical structure of land cover units (Figure 2).

While this structure of observations can initially appear cumbersome, it provides significant flexibility several different types of land cover attributes can be relevant to a single exemplar. The 19 Victorian land cover target classes were structured aggregations of the 40 land cover attributes (Table 2). Figure 2 clearly indicates the Victorian land cover classes represented in the Victorian Land Cover Layers 1985–2019. These are necessarily a subset of all those available and are designated with a red border, in one instance a completely new aggregation for Victorian purposes (i.e. horticulture / irrigated pasture or crop).

Table 2 Land cover attribute classes relevant to all epochs.
These attribute classes are modelled collectively within a single 'multi-objective' model.

Land Cover Class Attributes		
Native	land_vegetated_woody	native_shrublands
notNative	treed	woody_NotTreed_NotNative
Water	treed_native	land_vegetated_herb
water_deep	treed_native_roadside	land_vegetated_herb_wet
marine_grass	treed_native_paddock	herb_wet_NotNative
marine_bareground	treed_NotNative	herb_wet_Native
unvegetated	treed_NotNative_plantation	herb_wet_Native_perennial
Rock	plantation_pine	herb_wet_Native_occasional
Sand	plantation_BlueGum	land_vegetated_herb_dry
Built	orchard	herb_dry_crop
bareground	treed_NotNative_urban	grasslandpasture_native
Coal	windbreak	grasslandpasture_Not_native
disturbed	willows	
land_vegetated	woody_NotTreed	

Exemplar database

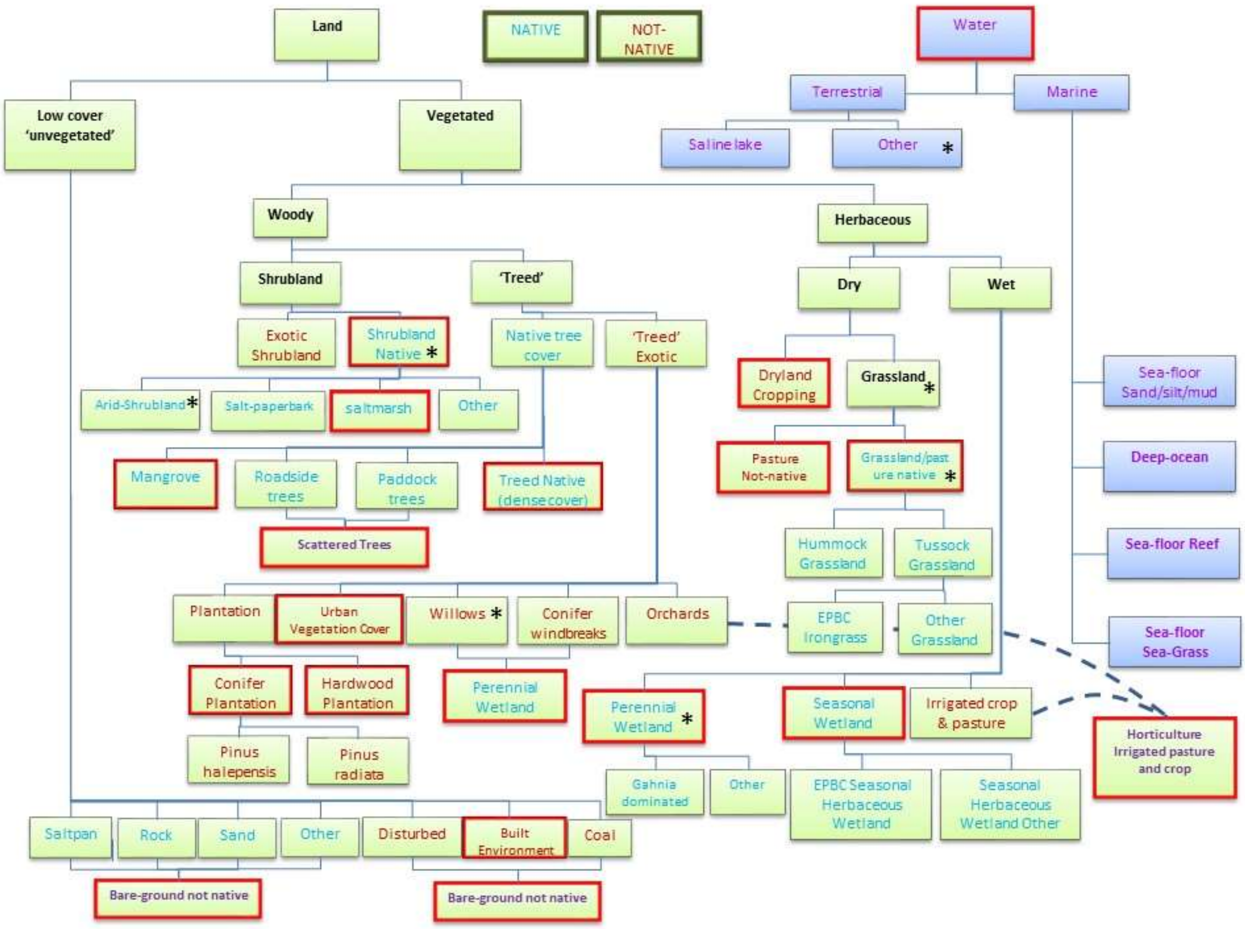
A land cover exemplar database was developed to collate the dependent data, and to support the inclusion of additional iterations of multi-temporal observations as these became available. The database structure was intentionally simple, where exemplars were time-stamped positions of land cover classes that were collated from various projects. These were collated into a single table called ExemplarLibrary. This table contains the class, position coordinates, data source, and the original identifier from that source and usage flags. Two additional tables comprise the complete database:

ExemplarClasses is a lookup table that lists all classes (i.e. attributes) that have thus far been added to the database. New projects may be added that extend the list. Generalizing classes can be applied easily prior to data extraction.

ExemplarSources table records the source project for the exemplar data, as well as its general definition date. This date may be used to restrict queries of the linked exemplar to timeframes that may be relevant to any specific project.

Figure 2 Hierarchy of land cover class attributes.

Hierarchy stems from a 'decision tree' logic for terrestrial land cover (green) and water (purple text). Aqua-colored text indicates land cover categories related to native vegetation types; exotic terrestrial land cover types indicated by red text. Red border on attribute boxes indicates the nineteen target land cover classes for Victorian data products.



NB: Several land cover class assignations / attributes relate specifically to South Australia, and are not present in Victoria, including EPBC Irongrass, Kangaroo Island Sheoak, Peppermint/Box, and *Pinus halepensis.*

Independent data

Independent data are a series of datasets necessary for establishing relationships to the dependent data (i.e. land cover observations). Producing spatial expressions of the modelled relationship(s) between the dependent and independent data requires that the latter have a coincident coverage across the complete study area for all independent datasets. The independent data for this project consisted entirely of remote-sensed satellite data that could be obtained at little or no cost from the public domain. The primary source for this data is from the Landsat Thematic Mapper (TM) satellite series that have been in operation for several decades delivering spectral data across six bands (Table 3).

Table 3 Landsat Thematic Mapper bands.

Landsat 7 Bands
Band 1 - Blue
Band 2 – Green
Band 3 - Red
Band 4 – Near Infrared (NIR)
Band 5 – Shortwave Infrared (SWIR - short) 1
Band 7 – Shortwave Infrared (SWIR - long) 2

A total of 72 input independent data layers were created for model development (Tables 4 - 6, with additional information provided in Appendix 2). For the median Landsat images and their ten indices, the correct image relevant to the exemplar date was used to supply the values. The remaining layers (Alos25_HV_min, etc.) were extracted for all exemplars.

Table 4 Independent data layers used for model development.
Full details are provided in Appendix 2. ALOS refers to synthetic aperture radar data (PALSAR) obtained from the Japanese Daichi 2 Advanced Land Observation Satellite.

Autumn	Winter	Spring	Summer	Min/Max & ALOS data
Median_Autumn_Band1	Median_Winter_Band1	Median_Spring_Band1	Median_Summer_Band1	MinMaxVic_Band1
Median_Autumn_Band2	Median_Winter_Band2	Median_Spring_Band2	Median_Summer_Band2	MinMaxVic_Band2
Median_Autumn_Band3	Median_Winter_Band3	Median_Spring_Band3	Median_Summer_Band3	MinMaxVic_Band3
Median_Autumn_Band4	Median_Winter_Band4	Median_Spring_Band4	Median_Summer_Band4	MinMaxVic_Band4
Median_Autumn_Band5	Median_Winter_Band5	Median_Spring_Band5	Median_Summer_Band5	MinMaxVic_Band5
Median_Autumn_Band7	Median_Winter_Band7	Median_Spring_Band7	Median_Summer_Band7	MinMaxVic_Band7
Median_Autumn_NDVI	Median_Winter_NDVI	Median_Spring_NDVI	Median_Summer_NDVI	Vic_NDVI_Delta1000
Median_Autumn_NDMI	Median_Winter_NDMI	Median_Spring_NDMI	Median_Summer_NDMI	WaterFromSpace
Median_Autumn_NDSI	Median_Winter_NDSI	Median_Spring_NDSI	Median_Summer_NDSI	Alos25_HV_Min
Median_Autumn_SLAVI	Median_Winter_SLAVI	Median_Spring_SLAVI	Median_Summer_SLAVI	Alos25_HV_Median
Median_Autumn_SATVI	Median_Winter_SATVI	Median_Spring_SATVI	Median_Summer_SATVI	Alos25_HV_Max
Median_Autumn_EVI	Median_Winter_EVI	Median_Spring_EVI	Median_Summer_EVI	AlosRatio
Median_Autumn_NDWI	Median_Winter_NDWI	Median_Spring_NDWI	Median_Summer_NDWI	
Median_Autumn_NDBR	Median_Winter_NDBR	Median_Spring_NDBR	Median_Summer_NDBR	
Median_Autumn_BATHY	Median_Winter_BATHY	Median_Spring_BATHY	Median_Summer_BATHY	

Landsat reflectance data – data-cube summaries, season, epochs

The majority of the independent data were derived from Landsat, either as summarized data, indices, or as published derived products (e.g. Water Observations from Space, Geoscience Australia). The Landsat platform supports a series of passive sensors that sample reflectance data from the Earth's surface across a wide swath of bandwidths. The nature and history of Landsat TM lends itself well to this modelling task, as this meso-scaled data (25 m pixels) has a 30 year+ library of available legacy data. Landsat TM data has had a long history of being used for land cover analyses over several decades, including the production of many synthetic datasets that are currently in use in the environmental sciences and natural resource management (NRM) domains.

The primary reflective satellite image data consisted of temporal Landsat median images covering the entire state. The Landsat image stack consisted of seven summarized epochs: (1987-90, 1990-95, 1995-00, 2000-05, 2005-10, 2010-15, 2015-19), and each epoch was divided into 4 overlapping seasons (Spring = Aug-Dec, Summer = Nov-Mar, Autumn = Feb-Jun, Winter = May-Oct). All available Landsat tiles (subsequent to error detection and cloud removal), were used to calculate a median image relating to each season in each epoch. In general, approximately 40 images contributed to each summarized image, although this did vary on a pixel-by-pixel basis within each summary image due to cloud removal, errors, etc. Summary images for first and last epochs (1985-90 and 2015-19) necessarily had fewer images due to the bounds of imagery that were available. Ten indices were also calculated for each epoch from the six native bands contained within each seasonal median image, and are listed in Table 5.

Table 5 Derivation of 'passive' indices used as independent data.

Pixel level values for indices were derived from the primary (median), and Landsat bands (displayed below).

Abbreviation	Index	Calculation from Landsat Bands
NDVI	*Normalised Difference Vegetation Index*	= (NIR-RED)/(NIR+RED)
NDMI	*Normalised Difference Moisture Index*	= (NIR-SWIR)/(NIR+SWIR)
NDSI	*Normalised Difference Snow Index*	= (RED-SWIR)/(RED+SWIR)
SLAVI	*Specific Leaf Area Vegetation Index*	= (((SWIR-RED)/(SWIR-RED+0.5))*1.5)-(LWIR/2) (raw values scaled to 1 prior)
SATVI	*Soil-adjusted Total Vegetation Index*	= NIR/(RED+SWIR)
EVI	*Enhanced Vegetation Index*	= (NIR-RED)/(NIR+(6*RED) -(7.5*BLUE) + 1)
NDWI	*Normalised Difference Water Index*	= (NIR-GREEN)/(NIR+GREEN)
NBR	*Normalised Burn Ratio*	= (NIR-LWIR)/(NIR+LWIR)
BATHY	*Bathymetric Index*	=((ln(BLUE-NIR)/ln(GREEN-NIR)) * -0.5 +0.5) *3E4 = 1500*(1-Ratio)
Vic_NDVI_Delta1000	*NDVI Delta*	= Mean NDVI taken over 5x5 pixels was differenced with the NDVI value at the centre of the pixel, with the calculated value is assigned to that pixel

Supporting Imagery

The other remote-sensed data source used in developing this landcover time series was radar data sourced from the Advanced Land Observing Satellite (ALOS 2) satellite from Japan. This platform supports a synthetic aperture radar sensor that provides indicative values for height above ground level, which can provide some structural information. In contrast to the Landsat platform that captures reflectance data, this is an 'active' sensor where the satellite is both the source as well as the receiver of radar data. Table 6 outlines the supporting indices used in model development.

Table 6 Additional Indices used as independent data

Independent Data Layer	Description
Alos25_HV_Min	Alos25_HV minimum value from each of the 2007 to 2010* statewide images
Alos25_HV_Median	Alos25_HV median value from each of the 2007 to 2010* statewide images
Alos25_HV_Max	Alos25_HV maximum value from each of the 2007 to 2010* statewide images
AlosRatio	A ratio of the Alos HV value as a pixel with the 5x5 pixel mean value around that pixel. This is similar to an edge-detect algorithm
MinMaxVic_Band1-Band7	Minimum Summer Landsat band values for all epochs divided by the maximum summer band values for all epochs. NB: this measure is epoch-independent, and was used to flag pixels that expressed massive fluctuations across epochs
WaterFromSpace	'Water Observations From Space' is a publicly available data layer which details the percent of time a Landsat pixel has been detected with water present across Australia. (Geoscience Australia - http://www.ga.gov.au/scientific-topics/hazards/flood/wofs)

* the period for which ALOS-PALSAR data was freely available

Model Development

Data Extraction and modelling

Exemplars for each epoch were extracted from the exemplar database for the full 40 classes relevant for each epoch, and were used as training data for modelling. Where an exemplar class was unlikely to change it was extracted for every epoch (e.g. classes such as rock, etc., or where classes were verified across epochs by examining of the terminal images). Other data extraction related to the specific epoch of the observation. Seven datasets of exemplars were compiled relating to each epoch, and these were used as the basis of the extraction of independent data (i.e. Landsat data, indices, etc.).

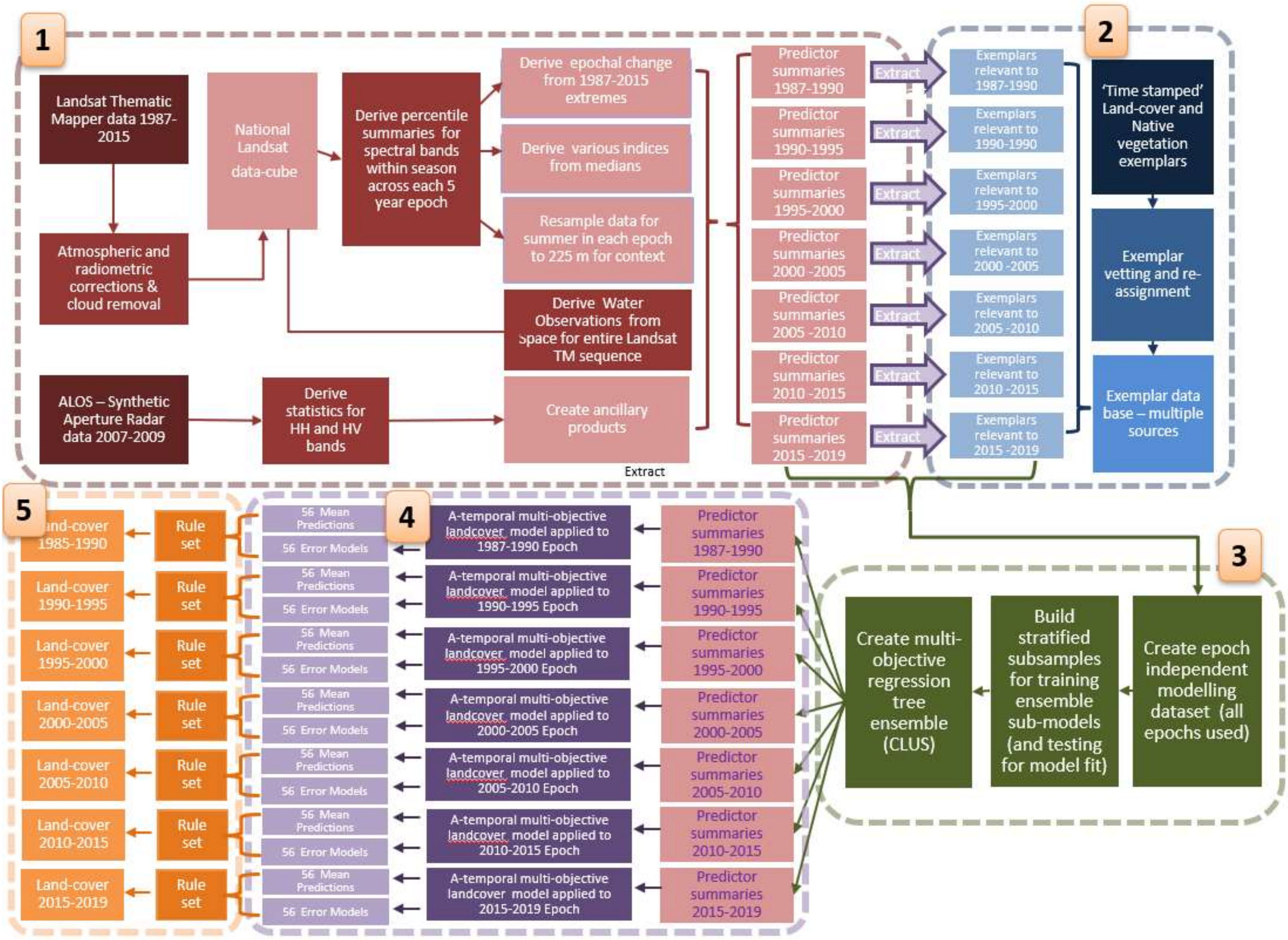
These seven independent data extractions were then concatenated to form a single model dataset covering all epochs to develop a single Random Forest Machine Learning model. There were several reasons for generating a single model. First, this approach provided a generic model rather than several different variants, and this was considered appropriate given the high stability of Landsat imagery over time. This approach also allowed for stability and consistency of modelled outputs (i.e. mapped products) across all epochs. Furthermore, this approach greatly simplifies the management of models and their outputs.

The single model was developed using *Clus* software, and was produced from 20 iterative ensembles of Random Forest models (Demšar *et al*. 2006; Kocev *et al* 2007). Each separate Random Forest was created from a randomised bag of observations that were stratified across the classes. The complete workflow of the modelling process is categorized into five major steps from the compilation of remote-sensed and exemplar data management and feature extraction, to stratified ensemble regression tree modelling, to the final output of seven sequential mapped land cover surfaces. This is displayed graphically in Figure 3.

Figure 3 Model building and application work-flow.

This chart is best interpreted sequentially from the upper left (1) to lower left quadrants (5).

Red regions and symbols display the steps relevant to the independent data compilation. Blue regions and symbols relate to steps relevant to the dependent data compilation (i.e. land cover exemplars). Green regions and symbols relate to the model development processes. Purple regions and symbols denote steps relevant model application (i.e. spatial expression of mapped outputs. Orange regions and symbols indicate steps relevant to model post-processing.



4 Results

Landcover attribute models

Models for each of the landcover attributes were spatially expressed into a single binary interleaved (BIL) files for each epoch consisting of 40 coincident layers where each interleaved layer for each pixel expressed a likelihood of that pixel being predicted any of the forty land cover attribute classes. Seven sequential models were developed that relate to each temporal epoch.

Model Fit

The overall levels of model fit for each of the landcover attributes within the general model was explored using correlations. The results display a reasonably high level of fit between predicted classes and those observed and provided for the model development (Table 7).

Table 7 Model fit for each of the 40 land cover attributes

Class	Correlation	R^2
bareground	0.844	0.713
Built	0.853	0.728
Coal	0.991	0.982
Disturbed	0.677	0.459
Grasslandpasture_native	0.871	0.758
Grasslandpasture_Not_native	0.855	0.732
herb_dry_crop	0.880	0.775
herb_wet_Native	0.781	0.609
herb_wet_Native_occasional	0.735	0.540
herb_wet_Native_perennial	0.778	0.606
herb_wet_NotNative	0.856	0.733
land_vegetated	0.909	0.825
land_vegetated_herb	0.947	0.896
land_vegetated_herb_dry	0.880	0.775
land_vegetated_herb_wet	0.815	0.665
land_vegetated_woody	0.944	0.892
marine_bareground	0.884	0.782
marine_grass	0.919	0.844
Native	0.898	0.807
native_shrublands	0.812	0.659
NotNative	0.900	0.809
Orchard	0.846	0.716
plantation_BlueGum	0.903	0.815
plantation_pine	0.909	0.826

Class	Correlation	R^2
Rock	0.800	0.640
Sand	0.706	0.499
Treed	0.941	0.885
treed_native	0.939	0.882
treed_native_paddock	0.621	0.386
treed_native_roadside	0.616	0.379
treed_NotNative	0.873	0.762
treed_NotNative_plantation	0.924	0.854
treed_NotNative_urban	0.681	0.464
unvegetated	0.869	0.754
Water	0.968	0.937
water_deep	0.897	0.805
Willows	0.595	0.354
Windbreak	0.625	0.391
woody_NotTreed	0.812	0.659
woody_NotTreed_NotNative	0.692	0.479

Classified Landcover model outputs

The seven independent stacks of 40 binary interleaved (BIL) files relating to each epoch constituted the primary data outputs. These probabilistic models of land cover attributes were then passed through a classifier module to process the stack of 40 probabilities relevant to each pixel into a form representing the most likely of the nineteen Victorian landcover classes that is likely to be present at each pixel.

This process allowed for some attribute classes to be subsumed within a single target class (e.g. marine_grass, marine_bareground and water_deep classes were merged to a single class termed ‘Water’). The output of this process was a series of seven raster layers relating to each individual epoch with pixel level representations of 19 landcover classes across Victoria.

An additional analysis was undertaken to comprehend the combinations of landcover classes that were most likely to present errors. The results of this analysis are displayed in a large confusion matrix of mis-classifications of pixels from within the final land cover grids (Appendix 3), indicating the landcover classes that were more difficult to reliably predict.

Post-processing of probability surfaces to obtain ‘most likely’ land-cover class

Temporal Filtering - rules

Up until this point in the production process, all epochal land cover grids had been produced effectively independently from each other. Whilst they used a commonly derived Random Forest model for all the landcover classes along with some common grids, the model for each epoch was based on the median Landsat reflectance data for the relevant specific epoch. In this way, each epoch was a separate entity, and only partially related to the data / information in other epochs.

From this perspective, apart from the initial generic model there were no clear reference points between any epoch that would support stability or cross-referencing across these epochs over time. This temporal or sequential referencing is essential to ensuring the production of stable and consistent data products. More specifically, the scenario of providing seven semi-independently derived outputs meant that relatively small differentials in predicted values for a landcover classes within a single epoch could cause the pixel to 'jump about' or 'flip' during the 'most-likely' final class assignation, despite the models being developed from very similar median imagery data across the epochs. This had the potential to add considerable instability when the final epoch layers were viewed across time, and particularly when appraised at finer local scales. For example, the classes treed_native_roadside, treed_native_paddock and treed_native are related entities, and slight differences in median imagery between epochs could cause inconsistent classification for the same pixel over time, despite there being little disturbance or change.

To ameliorate this prospect a temporal filtering process was devised to address potential inconsistencies in the classification outputs across the epochs. This process improved the consistency of the class models, and addressed specific problems in the outputs, particularly where the median imagery for the first and last epochs was relatively noisy due to limited data availability. The process for conducting this temporal filtering was encoded within the *ApplyPrediction* software, and a summary of the logic of the process is outlined below (Table 8). Each pixel from the classified outputs from the seven epochs were treated with the temporal filtering process.

Table 8 Logic of temporal filtering and class correction encoded within eight sequential tests

Step	*Task*
1	Is the output class consistent across all epochs? If so, do nothing.
2	If inconsistent across all epochs, (Else) determine the majority class (**MajClass**), the first epoch (**FirstClass**), last epoch (**LastClass**), second last epoch classes (2ndClass) and the sum of the Native probability across all epochs (**SumNative**).
3	Fix sudden switch to native after being predominantly non-native. If **MajClass** is non-native and **LastClass** or 2ndClass native, make those equal to the **MajClass**.
4	Similarly, fix first epoch being non-native if is native in later epochs. **If MajClass** is native and **FirstClass** is not native and **SumNative** > 420 (out of possible 600), make **FirstClass** = **MajClass**.
5	Check for native herb wetland at start and definite trend of being farmland. If **MajClas**s = Farmland and **StartClass** = herb_wet_Native or herb_wet_Native_perennial or herb_wet_Native_occasional and **SumNative** < 180 then **FirstClass** = **MajClass**.
6	Check for blinking on and off between start and end. If **StartClass** = **MajClass** and **LastClass** = **MajClass** and **MajClass** count = 5 and **MajClass** = **conifer plantation**, then all epochs = **MajClass**.
7	Check for consistency where the majority class just misses out in any single epoch. If **MajClass** count >= 3 and **MajClass** probability in this epoch is >= 0.5 output class probability, then output class = **MajClass**.
8	Finally check against a native forest mask for the large areas of forests and eliminate non-native trees that might speckle in and out of these areas. If in mask area and outclass = **ShrubNonNative** or **Pine** or **Bluegum**, select highest probability of classes **NativeShrubs**, **NativeTrees** and **NaturalLowCover** and assign that class.

Spatial Filtering - rules

The next process in creating the most likely land-cover class was to apply a spatially explicit 'cover-class reassignment surface'. These reassignments typically involved the detection of impossible or highly unlikely cover classes within a particularly cadastral unit or context, and was handled by reassigning the cover to the most likely alternative class. This process was applied to the outputs from the temporal filtering, e.g. very small instances of dryland crop in the Mallee region National Parks, or occasional instances of urban pixels within uninhabited coastal reserves.

Outputs of final Landcover Grids

The final maps consisted of seven grids in geographic projection of 40000 x 28000 pixels at a 0.00025 degree (approx. 25m) resolution, where each grid displayed the 19 target land cover classes across Victoria. Versions of these raster datasets were also reprojected into the 'VicGrid' format and supplied to Biodiversity Division (26520 x 28420 pixels at 25m resolution).

Additionally, there are seven continuous grids providing the probability of the winning class at each pixel for each epoch. These data were also supplied with an additional seven-layer grid of the standard deviations of the probabilities of the winning class across all seven epochs, which provided a quick reference to regions of where large changes in land cover were detected. This grid listed the same value across all layers if the class was entirely consistent across all epochs. If for example, in the first two epochs the class of a pixel was Native Pasture and it subsequently switched to Dryland Cropping, then the first two layers contained the Native Pasture probability standard deviation as calculated across all seven epochs, and the last five layers of the data stack contained the Dryland Cropping probability standard deviation across all seven epochs.

Caveats and Limitations

As with any body of work there are limitations and caveats that apply to the delivered products. End users of the products need to be cognizant of these factors, as they influence the utility and reliability of these data products.

Purpose and intent: This research program and resulting data sets were devised to produce a series of land cover data products that were more informative for the purposes of conservation planning and management, to support NRM activities, and to improve the ability to reliably monitor native vegetation extent over time. This need arose because existing published datasets have generally been focused to other sectoral interests (e.g. agriculture, forestry, economic activity, etc.), and have had limitations in NRM applications. We anticipate that these multi-temporal land cover data sets will be more useful for NRM practitioners, however this needs to be evaluated over time. Furthermore, caution should be taken when considering these data for other purposes beyond the initial intent.

Remote-sensed data: Remotely-sensed data is complex, and requires significantly levels of correction and pre-processing prior to any analysis, including with respect to land cover. As with any complex data set, errors and inconsistencies can never be completely discounted. The data used here have been corrected and pre-processed by Geoscience Australia, and have been undergone consistency checks by ARI staff prior to model development. We are also aware that the quality and acuity of the median imagery does vary over time, and as mentioned in the text, the data with first and last epoch will be subject to more noise, as there are less primary images available to calculate median pixel values.

Modelled data outputs: The data products produced from this process are modelled outputs, and have not been physically mapped either on-ground, or by using other technologies such as aerial photographic interpretation. By definition, models will have errors, and while all attempts have been made minimize these model errors, and to indicate where these errors may manifest (e.g. frequency of mis-classifications displayed in Appendix 3), end users are advised these datasets are indicative.

Scattered trees: The physical dimensions of an isolated paddock tree will generally fall below that of a single (median) Landsat pixel. This means that a pixel representing scattered trees is unlikely to have clean reflectance signal, and the feature will be represented as a 'mixel' (mixed or noisy pixel). These noisy numeric data will impact on model performance for these features.

The scattered tree class may also be under-estimated in more mountainous areas, and conversely maybe potentially be over-estimated in more agricultural settings. Scattered trees that are small in stature or development, or have fine or thin foliage or form (e.g. *Allocasuarina* spp.), can be difficult to detect with imagery resolved to 25m pixels. As a result, detections may be unreliable, although these limitations have not been quantified, and were beyond the project scope.

Class maps: The final mapped products from this modelling process are land cover classes. The classes have been modelled on a large series of numeric datasets that have considerably more information content than the representation using classes.

Temporal issues: The RS data used in the development of these land cover products are summarized to five-year epochs. While some land cover classes are relatively static, other units can be highly dynamic on a seasonal basis, or transient over time. Examples of these land classes that are likely to be volatile within-epochs are cropping versus dryland pasture, inland water bodies, seasonally-herbaceous wetlands, pivot irrigation systems, etc. Dynamic changes in land cover within these and other classes cannot be reliably detected *within* epochs, and the 'noise' that these land cover changes bring to median images will influence the consistency across epochs over time.

Land cover / geographic regions that will be expected to have greater uncertainty: There are a small number of known regions that are likely to have higher levels of uncertainty in classification.

> *Grampians – Billywig region, and Delatite Arm* – Eildon region are both going through a longer-term conversion from Pine Plantation to Tree_Native.

> *Strzlecki Ranges* - This region can be difficult to distinguish between Tree_Native and hardwood plantation, and the results for this region should be considered with caution. This issue is best addressed by considering tenure and land-use history as well as land cover, however a more nuanced product of this nature is beyond the scope of this project.

5 Discussion

In this project we have developed multi-temporal series of landcover maps at 25-m resolution across Victoria spanning more than 30 years to quantify dynamic changes in the extent, arrangement and configuration of native vegetation. To achieve this, observations were made of forty landcover attributes relevant across all epochs (1985–2019). These were compiled into a database and used to extract spectral information from 72 data layers derived from remote sensed imagery mostly from Landsat Thematic Mapper imagery which was used to develop a multi-target, multi-temporal Random Forest model of landcover that was consistently operable across all epochs.

The resulting series of models was internally tested and validated within the modelling process. The landcover maps had an overall prediction accuracy of >90% with high levels of correlation between observed and predicted pixels for native vegetation ($r = 0.898$), non-native vegetation ($r = 0.90$) and water ($r = 0.968$), demonstrating the ability of the process to differentiate between these key landcover attributes. While lower correlation coefficients were observed for some patchy and scattered landcover types such as windbreaks and scattered trees, the overall products have a high level of reliability as a result of post-processing temporal filtering rules.

Dynamic landcover information is critical for prioritising conservation efforts and managing responses to environmental disturbances, not in isolation but within the context of other broadscale land use practices focused on agricultural production, urban expansion, etc., and how the spatial mix of these systems alter over time Many landcover products to date have been single snapshots and therefore have had limited capacity to support long-term dynamic monitoring and environmental reporting. Additionally, broad-scaled landcover methods are relatively rare at medium resolutions, particularly with the breadth of landcover classes that have been modelled in this project. Our methods and multi-temporal data products effectively contextualism the trajectories of dynamic ecological changes at a meso-scale.

In this project we have provided a framework and aligned data systems to deliver consistent landcover data products suitable for reporting on native vegetation extent and landcover units relevant to conservation planning and monitoring. This process for producing landcover maps is transparent, repeatable and readily updatable, with many potential applications in the field of conservation and environmental monitoring. Importantly, these landcover data products are internally consistent and can therefore be used appraising the longer-term dynamics and accounting of landcover changes. The availability of free global satellite data also provides the opportunity for these methods to be applied to other regions at relatively low cost.

References

Demšar, D., Džeroski, S., Larsen, T., Struyf, J., Axelsen, J., Bruus Pedersen, M., and Henning Krogh, P. (2006). Using multi-objective classification to model communities of soil microarthropods. *Ecological Modelling* **191**(1), p. 131-143. https://doi.org/10.1016/j.ecolmodel.2005.08.017.

Kocev D., Vens C., Struyf J., Džeroski 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

Willoughby, N., Thompson, D., Royal, M. and Miles, M. (2017). South Australian Land Cover Layers 1987–2015: An Introduction and Summary Statistics, DEWNR Technical report 2017/09, Government of South Australia, Department of Environment, Water and Natural Resources, Adelaide.

https://data.environment.sa.gov.au/Content/Publications/SA-Land-Cover-Layers-1987-2015-Technical-Summary.pdf

Appendices

Appendix 1 – Definitions of the target land cover classes relevant for Victorian datasets.

Land Cover Class	*Description*
Water	Persistent surface water either fresh or saline – includes rivers, lakes, dams, wetlands and the ocean
Bare-ground – Native	Environments that naturally have low to negligible vegetation cover such as coastal foredunes, saline lake-beds, claypans and rock-outcrops.
Bare-ground - Not Native	Persistent unvegetated areas that are the result of anthropogenic activity other than urban development such as mining.
Built Environment	Persistent unvegetated areas that are the result of commercial or industrial development.
Scattered Trees	Native trees scattered in paddocks and woodland along roadsides and streams.
Urban Vegetation Cover	The admixture of streets, houses and gardens that characterises much of the medium to low density urban landscape typical of Australian cities.
Grassland/Pasture-native	Grasslands and pastures that are predominantly composed of indigenous species grasses and/or low chenopod shrubs. Includes grasslands that have been 'derived' through the clearing of tree and/or shrub cover.
Pasture/Grassland-Not native	Herbaceous pastures that are predominantly composed of non-indigenous species.
Dryland cropping	Regions that are regularly cropped and are not irrigated.
Wetland perennial – native	Persistent, typically herbaceous cover comprised of native plant species that are tolerant of inundation or waterlogging.
Wetland seasonal / ephemeral - native	Seasonal or ephemeral, typically herbaceous cover comprised of native plant species that are tolerant of episodic inundation or waterlogging.
Treed - Native	Native tree cover
Shrub-land – Native	Native Shrub cover
Hardwood plantation	Tree plantations predominantly *Eucalyptus globulus*
Conifer plantation	Tree plantations principally *Pinus radiata*
Other exotic tree cover	Non-native tree-cover including conifer windbreaks, willows along streams and rivers and varied ornamental plantings.
Horticulture/irrigated pasture and crop	Regions of crop, pasture and parkland regularly subject to irrigation, particularly in dry months.
Saltmarsh	Intertidal wetlands supporting native vegetation that are not mangroves
Mangrove	Intertidal native vegetation supporting *Avicennia marina*

Appendix 2 - Independent variables used in the generation of the land cover layers

Variable	Satellite	Pixel resolution	Season used	Statistics
Landsat B1	Landsat	25m	Summer, Spring, Autumn, Winter	25th, 50th, 75th percentiles for each epoch
Landsat B2	Landsat	25m	Summer, Spring, Autumn, Winter	25th, 50th, 75th percentiles for each epoch
Landsat B3	Landsat	25m	Summer, Spring, Autumn, Winter	25th, 50th, 75th percentiles for each epoch
Landsat B4	Landsat	25m	Summer, Spring, Autumn, Winter	25th, 50th, 75th percentiles for each epoch
Landsat B5	Landsat	25m	Summer, Spring, Autumn, Winter	25th, 50th, 75th percentiles for each epoch
Landsat B7	Landsat	25m	Summer, Spring, Autumn, Winter	25th, 50th, 75th percentiles for each epoch
Enhanced Vegetation Index	Landsat	25m	Summer, Spring, Autumn, Winter	Median for each epoch
Normalised Difference Moisture Index	Landsat	25m	Summer, Spring, Autumn, Winter	Median for each epoch
Normalised Difference Soil Index	Landsat	25m	Summer, Spring, Autumn, Winter	Median for each epoch
Normalised Difference Vegetation Index	Landsat	25m	Summer, Spring, Autumn, Winter	Median for each epoch
Soil Adjusted Total Vegetation Index	Landsat	25m	Summer, Spring, Autumn, Winter	Median for each epoch
Specific Leaf Area Vegetation Index	Landsat	25m	Summer, Spring, Autumn, Winter	Median for each epoch
Normalised Difference Burn Ratio	Landsat	25m	Summer, Spring, Autumn, Winter	Median for each epoch
Normalised Difference Wetness Index	Landsat	25m	Summer, Spring, Autumn, Winter	Median for each epoch
Spectral Bathymetry	Landsat	NA	Summer	Median for each epoch
Horizontal Transmit - Vertical Receive Polarisation (HV)	ALOS PALSAR (L-band)	25m	N/A	Minimum, Maximum, Median for 2005-2010 epoch
Horizontal Transmit - Vertical Receive Polarisation (HV)	ALOS PALSAR (L-band)	25m	N/A	Minimum, Maximum, Median for 2005-2010 epoch
Water Observations from Space	Landsat	25m	N/A	% of images detecting water
Alos Ratio	ALOS PALSAR (L-band)	25m	N/A	N/A
Landsat B1	Landsat	225m	Summer	Median
Landsat B2	Landsat	225m	Summer	Median
Landsat B3	Landsat	225m	Summer	Median
Landsat B4	Landsat	225m	Summer	Median
Landsat B5	Landsat	225m	Summer	Median
Landsat B7	Landsat	225m	Summer	Median
NDVI_Delta	Landsat	25m	Summer	Each epoch

Band1_MaxDiff	Landsat	25m	Summer	Each epoch
Band2_MaxDiff	Landsat	25m	Summer	Each epoch
Band3_MaxDiff	Landsat	25m	Summer	Each epoch
Band4_MaxDiff	Landsat	25m	Summer	Each epoch
Band5_MaxDiff	Landsat	25m	Summer	Each epoch
Band7_MaxDiff	Landsat	25m	Summer	Each epoch
Band1_MinDiff	Landsat	25m	Summer	Each epoch
Band2_MinDiff	Landsat	25m	Summer	Each epoch
Band3_MinDiff	Landsat	25m	Summer	Each epoch
Band4_MinDiff	Landsat	25m	Summer	Each epoch
Band5_MinDiff	Landsat	25m	Summer	Each epoch
Band7_MinDiff	Landsat	25m	Summer	Each epoch

Appendix 3 – Confusion Matrix of final land cover classes. Numbers indicate the number of pixels mis-classified from observed (columns) to predicted (rows)

	Observed																		
Predicted	Built Environment	Bare Ground - Not Native	Dryland Cropping	Exotic Woody Cover	Hardwood Plantation	Horticulture/Irrigated pasture & crop	**Mangrove**	**Grassland/Pasture - Native**	**Scattered Trees**	**Shrubland - Native**	**Tree - Native**	**Bare Ground - Native**	Exotic Pasture	Conifer Plantation	**Saltmarsh**	Urban Vegetated	Water	**Wetland Perennial**	**Wetland Seasonal**
Built Environment	1003	94										43	14			1208			3
Bare Ground - Not Native	95	10394	15			4		4		1	3	686	211	1		262	1		6
Dryland Cropping	2	485	85825	11		484		2401	110	547	104	293	3432			17			141
Exotic Woody Cover		376	12	857	33	20		37	28	20	39	5	171	2		85	1	13	
Hardwood Plantation		23	4	4	7247	3		11	33	36	717		10	333		4		36	
Horticulture/Irrigated pasture & crop	11	228	485	52	18	14077		78	11	16	20	22	779	1		169	1	50	139
Mangrove							221	3							13		1	3	
Grassland/Pasture - Native	16	77	1765	5	1	81		37424	209	829	1206	539	1886	9		98	19	27	1671
Scattered Trees		1	32	35		2		224	910	82	240	10	147	1		115	2	1	21
Shrubland - Native	3	10	50	9		52	3	165	1	6580	531	57	21	1	24	10	1	76	21
Tree - Native	1	40	48	133	108	50	4	452	212	2649	180673	616	127	496	36	426	31	141	93
Bare Ground - Native	6	208	117			6		218		64	106	7989	51			15	44	3	67
Exotic Pasture	2	641	4759	498	130	1694		5466	596	29	123	22	95182	5	4	94	5	817	1286
Conifer Plantation		21			108						205			10496					
Saltmarsh							3	52							406		3	18	
Urban Vegetated	80	116	16	37	1	100		6	3	5	2	70	81	3		10815		1	9
Water			3			2	2	27	2	1	7	239	30		2		3241	24	113
Wetland Perennial			5	25	1	27		100		157	28		24	3	32	2	6	2490	51
Wetland Seasonal	1	5	152	29		53		2128	17	172	115	21	379	1	71	16	8	466	14629

delwp.vic.gov.au