

FRONTIER S I >

VEGETATION EXTENT MAPPING IN THE MURRAY–DARLING BASIN

PROJECT REPORT

21 MARCH 2024

Joshua Kelcey, NGIS

Caitlin Adams, FrontierSI

Rory Donnelly, NGIS

Fang Yuan, FrontierSI

SUMMARY	2
1 INTRODUCTION	3
2 BACKGROUND	3
2.1 Data sources	3
2.2 Machine learning approaches.....	4
2.3 Classification scale	4
3 DATA.....	4
3.1 Ground samples.....	4
3.2 Remotely sensed data	5
3.3 Climate and environmental data	6
3.4 Land cover data.....	7
4 METHODOLOGY.....	7
4.1 Context	7
4.2 Data selection.....	8
4.3 Species selection	9
4.4 Training data creation	10
4.5 Training and evaluation of the machine learning models	15
4.6 Generation of the Basin-wide map	17
4.7 Post processing	17
4.8 Change detection.....	19
5 RESULTS.....	20
5.1 Model performance scores on validation set	21
5.2 Post processing	25
5.3 Species mapping.....	26
5.4 Change detection demonstration	36
6 DISCUSSION	38
6.1 Model training and validation.....	38
6.2 Species mapping.....	39
6.3 Change detection.....	41
7 CONCLUSIONS	41
APPENDIX.....	42
BIBLIOGRAPHY	43

Summary

This report summarises the work and outcomes of the Basin Condition Monitoring Plan project 3.4 “Vegetation extent mapping in the Murray–Darling Basin”, conducted by FrontierSI and NGIS. The goal of this project was to develop a method for mapping key vegetation types in the Murray–Darling Basin using annual satellite imagery. In addition, the method was to be used to produce a vegetation extent map for 2022. The project also developed a preliminary change detection approach, which can be further developed by the Murray–Darling Basin Authority (MDBA) to assist with reporting on species extent change.

The work was prompted by the need to monitor the extent of vegetation every five years as part of reporting against key outcomes of the 2019 Basin-Wide Environmental Watering Strategy (BWS), as well as community interests via the Basin Condition Monitoring Plan. Specifically, the relevant key outcome of the BWS is “to maintain the current extent of forest and woodland vegetation, including approximately 360,000 ha of river red gum (*Eucalyptus camaldulensis*); 409,000 ha of black box (*Eucalyptus largiflorens*); 310,000 ha of coolibah (*Eucalyptus coolabah*).”

Existing work conducted by Cunningham et al. (2013) produced a map using a decadal composite of Landsat satellite imagery. With the advent of higher resolution and more frequent satellite imagery captures, as well as advances in machine learning and cloud computing, this project aimed to develop a new method that could produce annual vegetation extent maps.

The project successfully developed a machine learning model that achieved an overall accuracy of 80.3% on 14 relevant vegetation species, a significant advance on the three species accurately modelled by Cunningham et al. (2013). In addition, the new method produces both individual species likelihood layers and a final classification extent map, providing additional information to the MDBA about the level of model confidence in any given area. The likelihood maps can also be used as the basis for customised vegetation extent maps, which could show extent with different levels of model confidence or incorporate multiple existing land use and land cover datasets to better map true vegetation extent. The project delivered the likelihood maps and a classification extent map for 2022 and demonstrated a change detection approach by running the approach for a small area comparing classifications from 2021 to 2022.

The project made efficient use of existing ground survey data by taking satellite measurements from multiple years for each species presence data point. This more than doubled the number of training points for the machine learning approach and produced a far more accurate model than could be achieved by only using satellite imagery from a single year. This approach also had the advantage of including inter-annual variability in the distribution of model features, supporting the model to make stable predictions across different years.

Given the customisability of the approach developed, the project team recommend that the next steps include user consultation about how the likelihood layers can be best used by the MDBA and its stakeholders. This should include dedicated effort to source independent local species extent maps that can be used to further validate the machine learning predictions, as well as customise the post-processing steps used to convert the likelihood layers to a final classification. As part of the project, the project team has delivered the model and processing software to the MDBA, and key staff have been trained to run the process. Importantly, this will enable the MDBA to run the existing process annually, as well as conduct the recommended further development.

1 Introduction

The extent of native vegetation across the Murray-Darling Basin is a key indicator of the health of the wider ecosystem. Specifically, a key expected outcome of the 2019 BWS is “to maintain the current extent of forest and woodland vegetation, including approximately 360,000 ha of river red gum; 409,000 ha of black box; 310,000 ha of coolibah.” (Murray–Darling Basin Authority 2019). The Murray–Darling Basin Authority (MDBA) monitors the extent of these species every five years in accordance with the strategy and would benefit from a system for measuring the annual extent of native vegetation that could also identify change. Such a system could quantify the impact of positive, human-driven interventions such as environmental watering, as well as long-term pressures such as climate change.

To be useful, the system must be comprehensive, consistent, automated, and allow for updates as the science of vegetation monitoring advances. Currently, monitoring of native vegetation occurs across multiple jurisdictions, without comprehensive coverage of the Basin. It is impossible for analysts to consistently measure complex and connected changes across the ecosystem. For analysts to meaningfully compare the measured extent year-on-year, as well as every five years in accordance with the BWS, the system must use a consistent process to generate that extent, allowing long-term changes to be trusted as a measure of impact from interventions and pressures. As science advances, as it invariably does, the system should allow for simple incorporation of these advances through a modular design, and past datasets should be re-computed with the latest science. Finally, robust automation to produce the vegetation extent each year ensures consistency and enables analysts to spend valuable time understanding the impacts of interventions on vegetation extent, rather than on running manual processes requiring huge amounts of data handling that are subject to human error.

The MDBA holds an initial version of such a system, developed by Cunningham et al. (2013), which used freely-available satellite imagery from Landsat and machine learning classification to map the extent of multiple species, with three native vegetation species having sufficient accuracy for the final extent map: river red gum, black box, and coolabah. The purpose of this project was to develop a new system that addresses three key limitations of the existing system: the mapping of only three vegetation types, the decadal timeframe on which maps are produced, and the lack of an established process for consistent and reliable change detection between maps.

This report describes the relevant background, data, and methodology for the new annual mapping system, including a demonstration of how the system can be used for change detection. Results from the system are presented and discussed, with an accompanying collection of recommendations for future work based on the findings from this project.

2 Background

The aim of this section is to review current approaches for species mapping with machine learning, focusing on the use of remote sensing and environmental data as model features. The review encompasses international research published in the last 15 years, including two recent review papers on the topic (Fassnacht et al. 2016; Pu 2021). Two studies that have specifically looked at vegetation species mapping in the Murray–Darling Basin are included: Cunningham et al. (2013) used remote sensing data over a ten-year time period to develop a machine learning model that mapped the extent of three key species (river red gum, black box, and coolabah); Mokany et al. (2023) used environmental variables to develop a machine learning model to predict the probability that river red gum or lignum were present in the landscape, as well as models for other fauna.

2.1 Data sources

Remote sensing data is valuable for species mapping as it captures spectral reflectance, which may vary between species depending on their structural and chemical composition, as well as their phenology. Due to their large area coverage and frequent revisit time, data from medium resolution multispectral satellites are regularly used by studies on mapping vegetation with remote sensing. Based on their free availability, Landsat (Grossmann et al. 2010; Cunningham et al. 2013) and Sentinel-2 (Grabska, Frantz, and Ostapowicz 2020; Lim et al. 2020; Masemola, Cho, and Ramoelo 2020; Yang et al. 2021; Bolyn et al. 2022; Mohammadpour, Viegas, and Viegas 2022) are popular choices. A few studies have used hyperspectral data for distinguishing species (Shang and Chisholm 2014; Badola et al. 2021) as this offers better spectral resolution, which can be valuable for distinguishing species even within the same genus (Shang and Chisholm 2014). However, hyperspectral studies typically use sensors on aerial platforms (rather than satellite), which comes with the downside of high cost to capture and infrequent revisit due to that cost. Across both multispectral and hyperspectral studies, there was no clear common approach as to whether studies used a single data capture, or multiple captures as either a composite or a time series.

Many studies also found that they generated more accurate species classification models when including ancillary data in addition to multispectral or hyperspectral remote sensing data. Of the studies reviewed, the most common ancillary data added was a digital elevation model, which allowed the machine learning model to leverage information about the elevation, slope, and aspect of vegetation (Cunningham et al. 2013; Grabska, Frantz, and Ostapowicz 2020; Liang et al. 2020; Lim et

al. 2020). Two of the studies found that features from the digital elevation model ranked highly in model importance for machine learning (Cunningham et al. 2013; Liang et al. 2020). Along with multispectral data, vegetation studies also benefit from synthetic aperture radar (SAR) data, which captures information in the microwave region of the electromagnetic spectrum. Yang et al. (2021) found that the combination of Sentinel-2 multispectral and Sentinel-1 SAR data produced a higher classification accuracy when mapping rice than either Sentinel-1 or Sentinel-2 independently. Finally, some studies have also leveraged environmental data such as temperature, precipitation, and soil content; Grossmann et al. (2010) included these variables alongside Landsat imagery, whereas Mokany et al. (2023) did not include any remote sensing spectral data. However, the inclusion of environmental data does not appear to be a common choice, as it was not mentioned at all the recent review of vegetation mapping using remote sensing by Pu (2021).

2.2 Machine learning approaches

Machine learning has been widely used for vegetation species classification and mapping in the last 15 years. Of the studies reviewed, the most common machine learning approaches were Support Vector Machine (SVM) and Random Forest (RF), both of which are nonparametric classifiers. Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs) were not as commonly used, which may be related to the fact they are challenging to train, a disadvantage noted by both Fassnacht et al. (2016) and Pu (2021) in their reviews of the literature.

Many studies used RF without comparing it to another machine learning approach, indicating its general acceptance within the remote sensing classification field (Grossmann et al. 2010; Cunningham et al. 2013; Lim et al. 2020; Mohammadpour, Viegas, and Viegas 2022). Both Grabska, Frantz, and Ostapowicz (2020) and Badola et al. (2021) found that SVM gave slightly higher overall accuracy scores when classifying vegetation species, with Badola et al. (2021) finding that the difference between the overall accuracy scores was not statistically significant. This may suggest that both methods are equally suited to the task.

2.3 Classification scale

An important choice when classifying vegetation from remote sensing data is whether to factor in the properties of nearby pixels when assigning a classification to any given pixel. Machine learning approaches can be applied to both individual pixels (referred to as the pixel-based approach) or groups of like pixels (referred to as the object-based approach). In this review, it was found that vegetation species classification studies tended to only apply pixel-based approaches, and did not also apply an object-based approach for comparison (Grossmann et al. 2010; Cunningham et al. 2013; Grabska, Frantz, and Ostapowicz 2020; Lim et al. 2020; Masemola, Cho, and Ramoelo 2020; Badola et al. 2021; Mohammadpour, Viegas, and Viegas 2022; Mokany et al. 2023). Only two studies applied an object-based approach (Shang and Chisholm 2014; Liang et al. 2020). In the single study identified that compared both approaches for vegetation mapping, Yang et al. (2021) found that the object-based method produced a higher overall accuracy score than the pixel-based method.

While there is no clear consensus on which approach is better, each offers different advantages. An advantage of object-based approaches is that they naturally group like pixels and then extract textural features such as spatial or temporal summary statistics (e.g. Yang et al. 2021), which can then be used as input features for the machine learning approach. While not a natural element of pixel-based approaches, textural information can be incorporated by selecting a local window and applying the grey-level co-occurrence matrix (e.g. Lim et al. 2020; Mohammadpour, Viegas, and Viegas 2022). Object-based approaches offer an advantage over pixel-based approaches in that they assign the same classification label to all pixels in an object, producing a more consistent species map without misclassified pixels within uniform vegetated areas, an issue encountered by Cunningham et al. (2013) when developing a pixel-based species classification for the Murray-Darling Basin.

3 Data

The aim of this section is to introduce the datasets that were used during the project.

3.1 Ground samples

For this project, the ground samples are records of locations where a species was observed during ecological surveys, and the date the record was made. With the ground samples, the machine learning model can be trained to identify relationships between remote sensing observations of the recorded species. The sections below summarise the collection of databases used for the project. The number of samples extracted from each dataset is presented in Table 1, Section 4.2.2.

3.1.1 Queensland – WildNet

The Queensland WildNet database provides ecological survey data for over 21,000 species of flora and fauna, collated from government agencies, researchers, businesses, natural resource management bodies, and citizen science programs. The database is managed by the Queensland Department of Environment, Science and Innovation (DESI). This database was chosen as it provided a straightforward interface for extracting the required data. The project team attempted to extract data from the Queensland Biodiversity and Ecology Information System (QBEIS), but were unable to extract the full complement of data through the available API. This was investigated for several weeks to no success, and the project team then decided to pursue WildNet and other sources (see Section 3.1.2).

3.1.2 Queensland – Private correspondence

Due to challenges accessing sufficient records for Queensland through QBEIS, the MDBA project team contacted the Queensland DESI about additional vegetation data that had been collected for the MDBA but had not been stored in QBEIS due to insufficient metadata records. The Department provided two datasets:

- Data on tree stand condition, collected in 2018,
- Data on woody species demographics, collected over 2019 and 2020.

Together, these datasets added over 1000 additional records to the Queensland data extracted from WildNet.

3.1.3 New South Wales – BioNet Atlas

The New South Wales BioNet is a repository for biodiversity data managed by the New South Wales Department of Planning and Environment. The repository contains the BioNet Atlas, which includes data collections on species sightings and systematic flora and fauna data collection. The repository contains millions of records across New South Wales, making it a useful source for this project.

3.1.4 Victoria – Victorian Biodiversity Atlas

The Victorian Biodiversity Atlas is a collection of species observations sourced from government and non-government organisations across Victoria and is managed by the Victorian Department of Energy, Environment and Climate Action. Much of the data were collected to assess species presence or abundance, making it highly relevant to this project.

3.1.5 South Australia – NatureMaps

NatureMaps provides an access point for maps and geographic information about South Australia's natural resources, including fauna and flora surveys. The portal is managed by the South Australian Department for Environment and Water. NatureMaps contained many records within the Basin, making it useful for this project.

3.2 Remotely sensed data

In the context of this project, remotely sensed data are satellite observations of Australia. These observations measure the reflectance of light at different wavelengths of the electromagnetic spectrum. Species of vegetation can be distinguished by their general reflectance characteristics, as well as how their reflectance changes with the seasons and other variables. For example, evergreen trees may have a more consistent reflectance profile across the year compared to deciduous trees that lose their leaves in winter, which may lead to a satellite observing lower amounts of green light for deciduous trees at that time.

The project used two remotely sensed datasets: Sentinel-2 and Sentinel-1, as described below.

3.2.1 Sentinel-2

Sentinel-2 is an optical imaging mission that captures 13 spectral bands ranging from visible light to shortwave infrared, at resolutions up to 10 metres. The satellite has a revisit time over Australia of around five days, although the number of useable observations is reduced by cloud cover.

Optical observations are valuable for vegetation detection. Vegetation can be distinguished from other land cover types in optical observations because vegetation has higher reflectance of near-infrared light than soil and water, as shown in Figure 1. The seasonal behaviours of different vegetation species, such as yellowing when drying, can also be measured by optical

observations, making Sentinel-2 data useful for distinguishing species. This project used Sentinel-2 surface reflectance data hosted by Google Earth Engine.

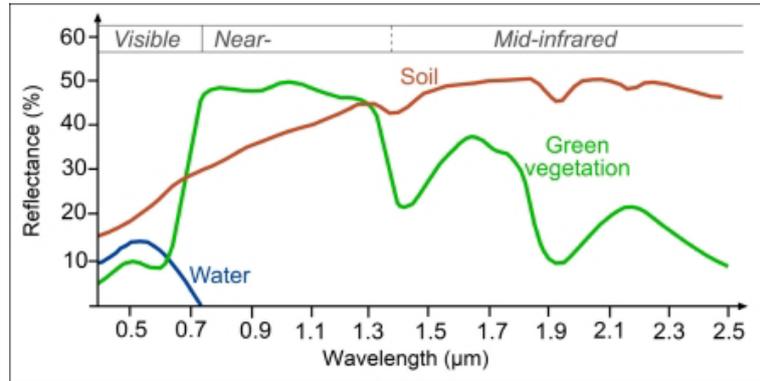


Figure 1. Spectral reflectance of soil, water, and green vegetation across the visible, near-infrared, and mid-infrared portions of the electromagnetic spectrum. The three land cover types are well distinguished by the amount of near-infrared reflectance (Source: Science Education through Earth Observation for High Schools).

3.2.2 Sentinel-1

Sentinel-1 is a SAR mission that captures reflectance in the C-band microwave range of the electromagnetic spectrum. As an active sensor, it emits light in this range, and then measures the amount of reflected light and its polarisation. Sentinel-1 has multiple capture modes which produce different resolution data; the relevant capture mode for this project is the Interferometric Wide Swath mode, which captures data at a resolution of 5 metres by 20 metres. The satellite has a revisit time over Australia of around 12 days. The C-band microwave range used by Sentinel-1 is unaffected by cloud cover.

Microwave reflectance provides information about the physical structure of vegetation. In particular, the C-band range captures information on the structure of canopies, which can help distinguish species with dense canopies from species with sparse canopies. The project used Sentinel-1 Ground Range Detected data from Google Earth Engine.

3.3 Climate and environmental data

Along with remotely sensed data, climate and environmental data (either derived from remote sensing observations or produced independently) are highly valuable. Along with climate, environmental conditions represent the main drivers of plant distribution and diversity. Within Australia, the strongest environmental predictors for a vegetation species are temperature, precipitation, soil characteristics and topographic heterogeneity. At a broad scale, these factors are driven by latitude and elevation, while local variation is driven by geology and topography.

The project used three additional datasets, which are described below.

3.3.1 SILO – Australian climate data

SILO is a gridded database of Australian climate data hosted by the Queensland DESI. The database provides daily meteorological datasets that can be used for biophysical modelling, such as rainfall, temperature, evaporation, solar radiation, and humidity. Data is sourced from the Australian Bureau of Meteorology and other suppliers.

For this project, climate data provides valuable contextual information for identifying different vegetation species, as different species will grow and thrive under different rainfall and temperature conditions. The project used a copy of the SILO database that is hosted as a public Google Earth Engine Asset by NGIS.

3.3.2 Soil and Landscape Grid of Australia (SLGA)

The Soil and Landscape grid of Australia (SLGA) is a comprehensive dataset of soil attributes across Australia. The dataset was developed by a partnership comprised of Australia's state, territory, and Commonwealth government agencies as well as the University of Sydney and is hosted by TERN. SLGA provides information on soil attributes such as clay, organic carbon, and soil depth, among many others.

Like climate data, data on soil properties are useful for distinguishing different vegetation species which grow and thrive in different soil types. The project used the SLGA data hosted by Google Earth Engine.

3.3.3 Shuttle Radar Topography Mission (SRTM)

The Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) is the result of an international research effort that obtained digital elevation models on a near-global scale (Farr, et al. 2007). Available worldwide, Geoscience Australia has developed products tailored to Australia, such as the Hydrologically Enforced Digital Elevation Model, which incorporates mapped stream lines.

Digital elevation models provide valuable information on the height, slope, and aspect of the terrain, which helps to identify riparian species that grow along river banks. The project used the Australian Hydrologically Enforced Digital Elevation Model data hosted by Google Earth Engine.

3.4 Land cover data

An important part of accurately mapping species extent is to exclude irrelevant land cover. Here, the satellite-derived land cover dataset used in this project is discussed.

3.4.1 Dynamic World

Dynamic World is a global, near-real-time land use/land cover dataset derived from Sentinel-2 (Brown, et al. 2022). For each Sentinel-2 scene, Dynamic World provides the probability that a pixel is completely covered by one of nine land use/land cover classes, as estimated by a machine learning model. A scene with the label of the highest probability class for each pixel is also provided.

For this project, land use/land cover data is used to mask out areas that are irrelevant for vegetation mapping in the Murray–Darling Basin, such as bare earth or built area. The Dynamic World dataset has an advantage over other land use/land cover datasets in that it provides the probabilistic classification as well as the highest-probability classification, which can allow us to construct more nuanced approaches for masking non-vegetated areas compared to datasets that only provide the highest-probability classification. One disadvantage is that it is a global dataset, with only ten validation tiles for Australia, meaning that the product is likely to be less accurate than a land use/land cover dataset that has been designed for Australia. The project used the Dynamic World data hosted by Google Earth Engine.

4 Methodology

This section of the report covers the project methodology, beginning with the project context before detailing the approaches used, covering:

- Data selection
- Species selection
- Training data creation
- Training and evaluation of the machine learning models
- Generation of the Basin-wide map
- Postprocessing
- Change detection

4.1 Context

Responding to recommendations from Cunningham et al. (2013), the project developed a method to map species on an annual basis. The project team had the advantage of leveraging Sentinel-2 data, which has higher resolution than the Landsat data used by Cunningham et al. (2013). In developing the new method, choices were made that mitigated challenges faced by Cunningham et al. (2013):

- Used both presence/absence and quadrat data when selecting records from state databases for machine learning. Cunningham et al. (2013) used quadrat data only, which limited the number of records that could be used for machine learning as these samples are more time intensive to collect than presence/absence, and consequently have fewer records in state databases.
- Used an object-based approach, as the pixel-based approach used by Cunningham et al. (2013) resulted in scattered presence of incorrect taxa among large vegetation communities, notably, scattered *Eucalyptus coolabah* in the Southern Basin.

4.2 Data selection

4.2.1 Remote sensing and ancillary data selection

Drawing from the datasets discussed in Section 3.2 and Section 3.3, 27 baseline features were selected for the model, categorized by what they measure: biophysical, structural, or environmental properties. All eight landcover classes from Dynamic World were selected for masking. The chosen features are summarised in Figure 2.

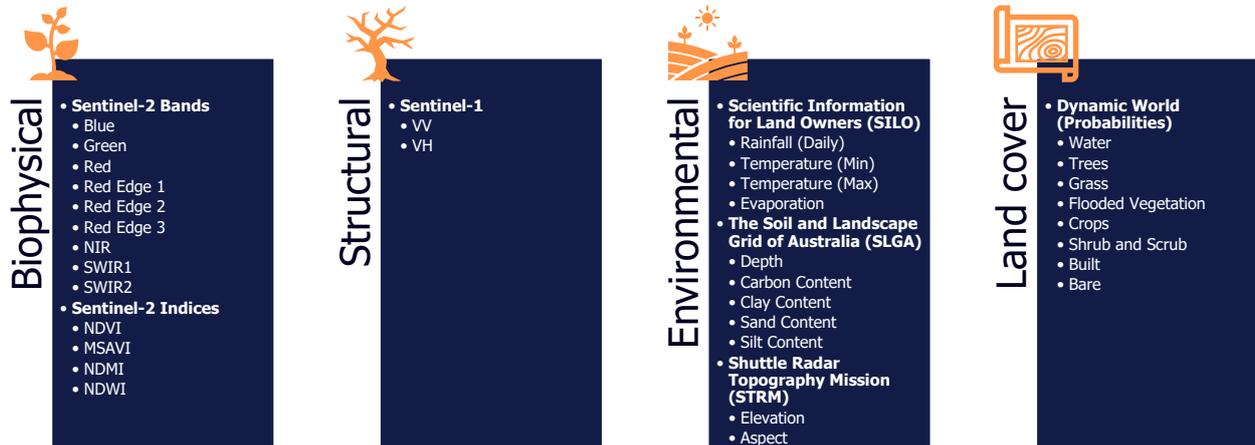


Figure 2. The datasets used in the project. The biophysical, structural, and environmental datasets were used in the modelling. The land cover datasets were used during post-processing.

The Sentinel-2 band indices used were the normalised difference vegetation index (NDVI) (Qi et al. 1994), the modified soil-adjusted vegetation index (MSAVI) (Qi et al. 1994), the normalised difference moisture index (NDMI)(Gao 1996), and the normalised difference water index (NDWI) (McFeeters 1996). The Sentinel-1 bands used were the vertical-vertical polarisation band (VV) and the vertical-horizontal polarisation band (VH).

4.2.2 Record selection

For this work, ground data were collected from databases covering Queensland, New South Wales, Victoria, and South Australia, as described in Section 3.2. Following the database extraction (which contained both flora and fauna), the databases were filtered down to the available flora records between 2018 and 2022 that had positional uncertainties of 1000 metres or less. For Queensland and South Australia, the records were restricted to those within the Murray–Darling Basin extent; for New South Wales and Victoria, all records were kept. The number of records found is shown in Table 1. The common attributes that were selected across all databases are shown in Appendix 1.

Table 1. Number of records extracted from databases for each state.

State	Database or source	Number of records without filtering (all flora and fauna)	Number of records (all species) after spatio-temporal filtering	Number of unique species after spatio-temporal filtering
Queensland	WildNet	48,630	3,364	455
	Stand Condition (2018)	1,938	436	3
	Woody Species Demographics (2019-2020)	900	900	18
New South Wales	BioNet Atlas	1,225,386	1,208,945	7,390
Victoria	Victorian Biodiversity Atlas	3,731,351	230,156	3,934
South Australia	NatureMaps	344,246	69,084	1,937
All States		5,352,451	1,512,885	13,737

4.3 Species selection

To determine which species could be successfully modelled, the project team developed the following species selection criteria in consultation with the MDBA team, guided by the relevant literature on remote sensing vegetation studies as described in Section 2:

- Can be detected by remotely sensed data (consistently present in the overstorey across the Basin).
- Must have a dominant presence in its local environment.
- Must be present in large extents across the Basin.
- Must be affected by water management (i.e., present in floodplains and wetlands or included in the Australian National Aquatic Ecosystems (ANAE) Classification System (Aquatic Ecosystems Task Group 2012)).
- Should have sufficient samples for machine learning (hundreds to thousands) or be sufficiently spectrally/environmentally distinct to be picked up by machine learning, even with a small number of samples.

The final point relied on data exploration, including examining how much variability there is within a given species across time. For example, it is challenging to use a single model to consistently map a species each year if the species properties alter drastically with the changing seasons and climate. This is discussed in more detail in Section 4.4.2.

The project team and the MDBA team selected 18 species for consideration during the model development stage. These are displayed in Table 2, along with the number of field records that were collected from the databases. These records underwent additional processing to produce generalised plots and finally a filtered set of machine learning samples.

Table 2. The 18 species selected for model development, with the number of field records, plots, and machine learning samples generated during the project. Machine learning samples differ from plots in that they have been sampled in multiple years and have had duplicates removed within segmentation objects. As such, they may be higher or lower than the number of plots. Species in red were removed from the final model as they could not be reliably modelled due to low sample count.

Scientific Name	Common Name	Number of field records (Section 4.2.2)	Number of plot records (Section 4.4.2: Refining label quality)	Number of machine learning samples after filtering (Section 4.4.2: Feature engineering)
<i>Acacia salicina</i>	Willow wattle	482	131	328
<i>Acacia stenophylla</i>	River cooba	1505	390	240
<i>Atriplex nummularia</i>	Old man saltbush	139	80	172
<i>Atriplex semibaccata</i>	Berry saltbush	1489	471	832
<i>Atriplex suberecta</i>	Lagoon saltbush	530	110	152
<i>Callitris glaucophylla</i>	White cypress pine	3601	1066	3300
<i>Casuarina cunninghamiana</i>	River oak	659	349	1108
<i>Chenopodium nitrariaceum</i>	Nitre goosefoot	634	222	132
<i>Duma florulenta</i>	Lignum	1506	650	536
<i>Eucalyptus albens</i>	White box	2629	1274	2223
<i>Eucalyptus blakelyi</i>	Blakely's red gum	2933	920	1845
<i>Eucalyptus camaldulensis</i>	River red gum	9117	3712	1510
<i>Eucalyptus coolabah</i>	Coolabah	1736	406	428
<i>Eucalyptus largiflorens</i>	Black box	3452	1734	668
<i>Eucalyptus melliodora</i>	Yellow box	3620	899	2413
<i>Eucalyptus microcarpa</i>	Grey box	2015	906	903
<i>Lycium ferocissimum</i>	African boxthorn	1839	1001	2632
<i>Nitraria billardierei</i>	Nitre bush	220	144	384
Total		38,108	14,465	19,806

4.4 Training data creation

4.4.1 Object-based segmentation

The project team used an object-based classification approach, implemented in Google Earth Engine. The object-based approach focuses on classifying collections of pixels with similar properties, leveraging their shared properties rather than treating every pixel individually. The input for the segmentation was the seasonal NDVI median values for 2022, such that the segmentation focused on identifying objects based on their seasonal vegetation presence.

The method chosen for segmentation was Simple Non-Iterative Clustering (SNIC) (Achanta and Susstrunk 2017), implemented in Google Earth Engine (Google 2023). SNIC works by dividing an image into regions of connected pixels, using a metric that accounts for distance in colour space and pixel space. Initial seeds are placed on a regular grid, then neighbouring pixels are checked and assigned to the closest centroid. The algorithm requires the user to set the segmentation size and compactness, as well as specifying which bands to use in the colour space. The following configuration settings were used:

- Segmentation size = 100
- Segmentation compactness = 0.2
- Segmentation bands = seasonal NDVI
 - Summer: 01/12/2021 – 28/02/2022
 - Autumn: 01/03/2022 – 31/0/2022
 - Winter: 01/06/2022 – 31/08/2022
 - Spring: 01/09/2022 – 30/11/2022

An example of the segmentation is shown in Figure 3.

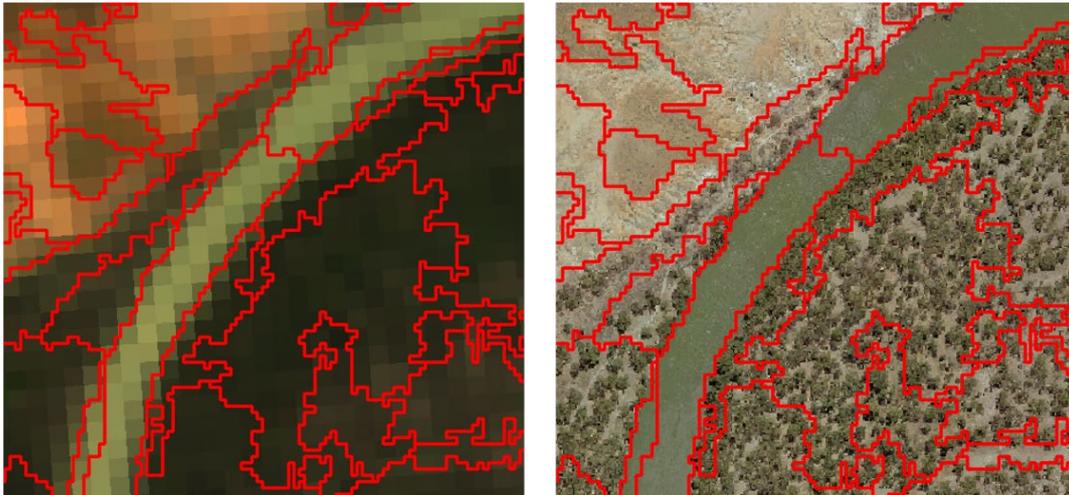


Figure 3. Objects (red polygons) identified from our initial segmentation of Sentinel-2 data overlaid on the 2022 Landsat 8 Geomedian from Digital Earth Australia (left) and Vicmap aerial imagery (right). This shows that the segmentation polygons delineate regions with different tree density, even when produced using 10 metre resolution data.

4.4.2 Sample creation

The training data for the machine learning model comprises many samples, each of which consists of a label and features. For this project, the labels are the ground samples, which provide known locations for the selected species. The features are the remote sensing, environmental, and climate datasets, which allow the machine learning model to identify relationships between the remote sensing observations, and the known species on the ground. The goal of this step is to provide the model with high-confidence labels and informative features, which are needed to produce an accurate machine learning model for measuring the extent of the chosen species across the Basin.

The following sections provide a summary of the steps taken to refine the labels and features before model training.

Refining label quality

Data quality, in this case referring to how reliably a record represents the true environment, is a key factor in developing a robust and accurate machine learning classifier. The project team undertook additional processing to collapse co-located records into plots and identify the dominant overstory species for each plot.

Step 1: Collapse samples into plots

The project team found that many of the selected records were co-located. This occurs when a field survey team records many measurements in a small area for operational efficiency. Our use of segmentation means that multiple records of different species may appear within a single object, which could cause model confusion. To address this, the records were first grouped into plots if they were collected within 50 metres and 30 days of each other. The plots could then be further analysed to select only the highest quality plots for model training.

Step 2: Assign confidence scores to plots

The highest quality plots are those that provide clear information for the machine learning model to learn from. If a plot contains samples of only a single species, it is likely that the area is dominated by this species and that the remote sensing features will be a good representation of the species, providing a clear example for the model to learn from. Alternatively, if there are many species within a plot, the remote sensing features will represent a mix of multiple species, which reduces the chance that the pixel will be accurately assigned a single species label. This is further compounded by the fact that species may be present at different heights within the plot, and that remote sensing datasets are most sensitive to the overstory, rarely picking up vegetation beneath canopy cover.

Based on these factors, the project team developed a method for assigning a confidence level to each plot, representing its suitability for model training (demonstrated as a flowchart in Figure 4). Species were first assigned to four height strata based on their typical height (Appendix 2). For plots with multiple species in the highest stratum, the model would likely receive information about all species, potentially causing confusion during model training if labelled as a single species. For plots where one species was found in the highest strata, and no species were found in the second highest strata, the model would likely receive information only about the species in the highest strata, with species below the second highest strata forming the understory. Finally, for plots where multiple species were found in the second highest strata, the model would likely receive some information about all species, potentially causing confusion during model training.

Based on this rationale, the following rules were designed to assign a confidence level to each plot:

- If there was only one species in the plot, that species was classed as dominant. The plot received a relative confidence score of 5.
- If there were multiple species present, the approach counted the number of species in each height strata, using the strata defined in Appendix 2. Then:
 - If there were multiple species in the highest strata, the plot was regarded as having no dominant species. The plot received a relative confidence score of 0 and was discarded.
 - If there was only one species in the highest strata, that species was classed as dominant. Then:
 - If there were no species in the second highest strata, the plot received a relative confidence score of 4.
 - If there were species in the second highest strata, the plot received a relative confidence score of 4 minus the number of species in the second highest strata. If there were four or more species, the subtraction produced a relative confidence score of zero.

Any plots with a confidence score of zero were discarded from the training dataset.

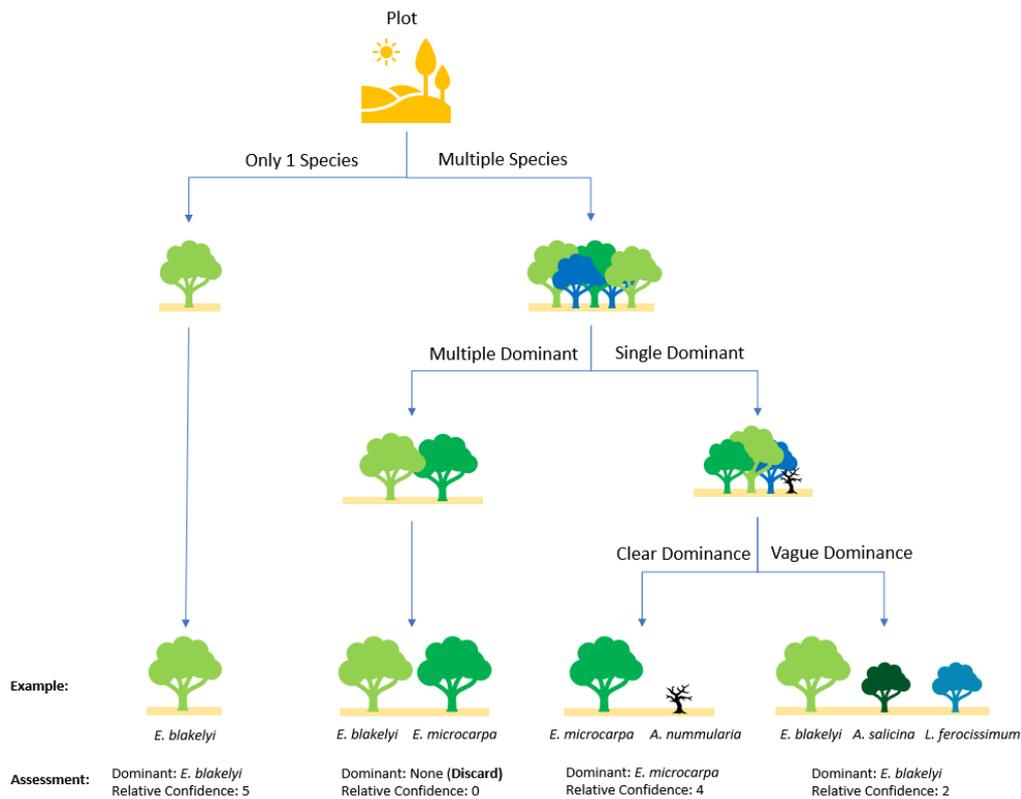


Figure 4. Diagram illustrating the rules-based method for determining dominant species and relative confidence.

Feature engineering

The biophysical, structural, and environmental features listed in Figure 2 were used in conjunction with the labels to train the machine learning model. While they can be used as is, feature engineering enables the generation of more relevant information for the model. Specifically, this included extracting additional temporal information and spatial information, as well as removing features that were likely to confuse the model. Feature engineering provides the model with more generalised information to learn from, reducing the chance of overfitting.

Step 1: Sample over multiple years

When going from species records to plots, the number of samples for machine learning decreased due to aggregation of multiple records into one plot. While this approach produces high confidence samples for machine learning, the project team found that there were then insufficient samples to produce a reasonable overall accuracy score when sampling from a single year. To increase the number of samples, features were measured for 2019, 2020, 2021, and 2022 for every sample. This assumed that the recorded species was present during all sample years; for example, if a field collection team logged the presence of *Casuarina cunninghamiana* in 2020, the method assumed that it was also present at that location in 2019, 2021, and 2022. This method had the added benefit of training the machine learning model to classify vegetation during either wet or dry years.

Step 2: Create additional features from statistical summaries

Statistical summaries of features can help to reduce variation across all samples within a species, making a more robust model. For this approach, both temporal and spatial statistical summaries were produced.

Temporal aggregations of Sentinel-1, Sentinel-2, and SILO data were performed to ensure the machine learning model features captured the temporal variation in our species of interest. Data was aggregated for four time periods within each year used: summer (01/12 – 28/02), autumn (01/03 – 31/05), winter (01/06 – 31/08), and spring (01/09 – 30/11), and the median value was calculated for each pixel.

The spatial statistics were generated by applying a collection of statistical measures on all pixels within the segmented object corresponding to each sample, namely the mean, standard deviation, minimum, maximum, kurtosis, and 10th, 25th, 50th, 75th, and 90th percentiles. This was done after the temporal statistic step, meaning these statistics were applied to the median seasonal values for each feature in each object, as shown in Figure 5.

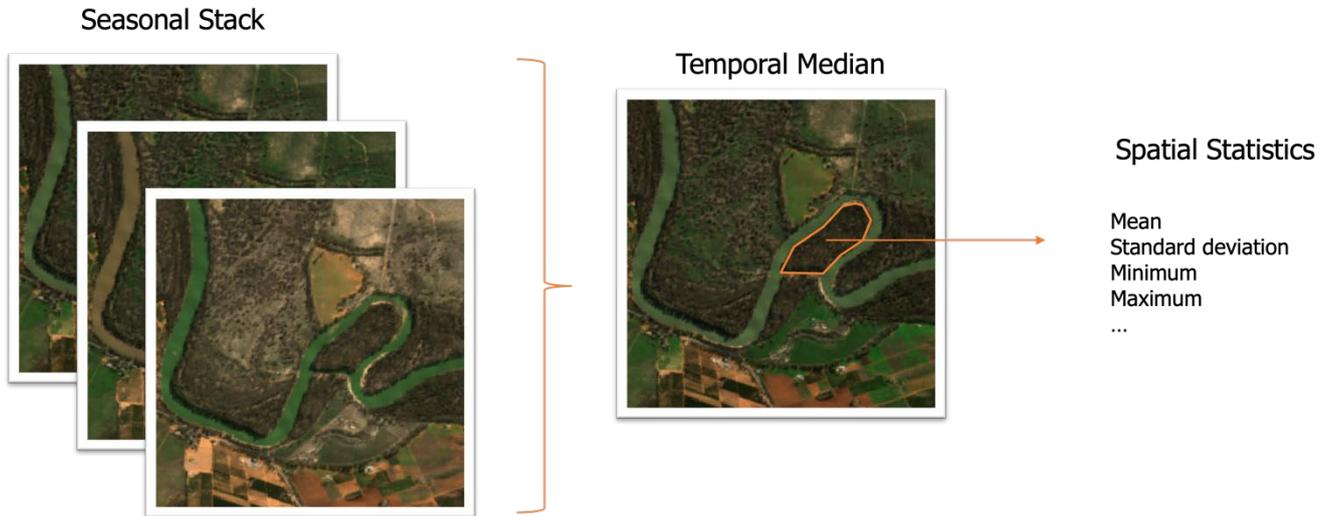


Figure 5. Temporal statistics are generated by grouping remote sensing features by season and taking the median. From the temporal median, spatial statistics are then calculated across the collection of feature pixels belonging to each segmentation object (orange polygon).

Step 3: Remove duplicate samples based on distance in feature space

After calculating all features, each sample occupies a location in the feature space. It is then possible to calculate the Euclidean distance between samples in this space: samples with similar feature values will be close together, and samples with dissimilar feature values will be far apart. If two samples of the same species are closely co-located in the feature space, it is sufficient to keep one sample and discard the other. This commonly occurs when multiple plots appear within the same segmentation object. If two plots of different species are closely co-located in the feature space, this is likely to cause confusion for the model, in which case both plots are removed. By removing duplicate and confusing samples, a higher-quality training dataset for the machine learning model was constructed.

Step 4: Remove features with high annual variability

One of the goals of this work was to produce a model that can be run across multiple years to facilitate change detection. The solution for this was to remove features that varied significantly depending on the year. For example, the project team found that the summer NDVI values for the samples tended toward low values in dry years, and high values in wet years. Based on this, all summer variables were removed, with the reasoning that the properties of vegetation in the summer are likely to vary significantly from year to year depending on the weather. This is demonstrated in Figure 6 – note the variability of the NDVI histograms in different years shown in the leftmost column compared with the relative stability of the interannual variation for the other seasons.

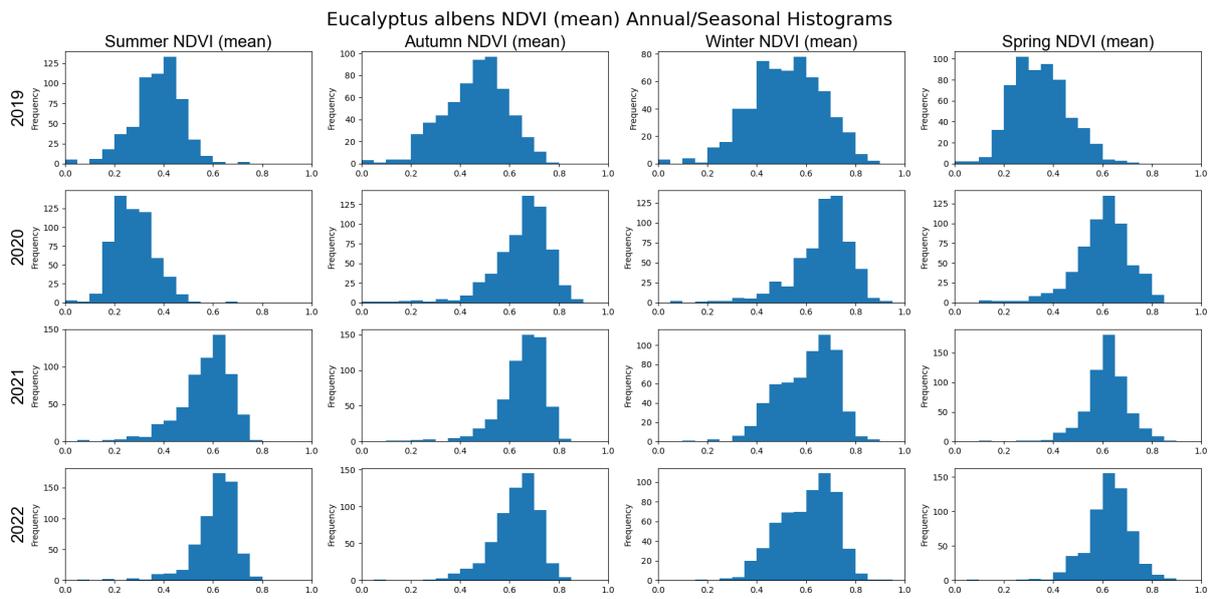


Figure 6. Comparison of the mean NDVI frequency distributions for *Eucalyptus albens* across four seasons (summer, autumn, winter, and spring) for four years (2019, 2020, 2021 and 2022).

Step 5: Remove any species with insufficient samples

Before commencing machine learning, the project team assessed whether the species had sufficiently consistent feature distributions each year from 2019 to 2022, noting that this encompassed both wet and dry periods in the Basin. It was found that species with fewer than 300 samples did not have sufficient volume to produce consistent feature distributions, making them difficult to accurately model. The mean NDVI feature distributions for *Atriplex suberecta* are shown in Figure 7 as an example. Based on this finding, four species were removed from consideration for modelling: *Acacia stenophylla* (240 samples), *Atriplex nummularia* (172 samples), *Chenopodium nitrariaceum* (152 samples), and *Atriplex suberecta* (132 samples).

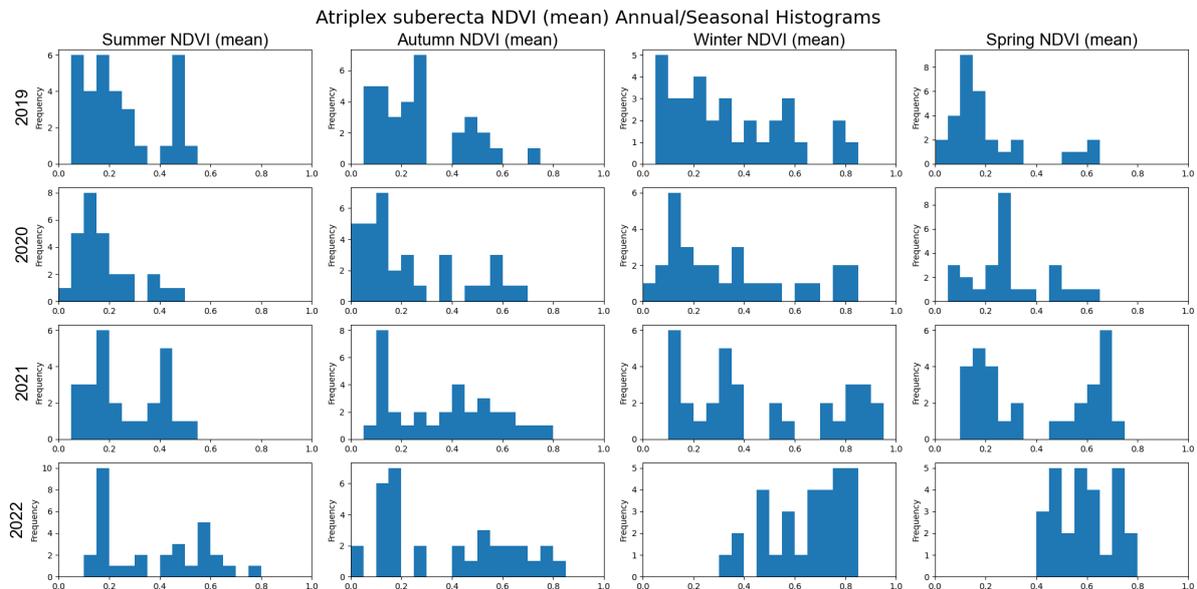


Figure 7. Comparison of the mean NDVI frequency distributions for *Atriplex suberecta* across four seasons (summer, autumn, winter, and spring) for four years (2019, 2020, 2021 and 2022).

Step 6: Remove correlated features

Producing the temporal and spatial statistics creates many features; of these, some may provide similar information to the machine learning model. It is possible to improve model efficiency by removing correlated features, thereby reducing the number of features that need to be calculated for training and prediction. Removing correlated features also reduces the risk of overfitting and reduces computational expense. For any features that were either positively or negatively correlated above an absolute correlation score of 0.99, the second feature was removed.

Step 7: Split samples into training and validation sets

Following machine learning common practice, the samples were divided into two sets, with 80% being used to train the model, and 20% used to validate the model. This allowed reporting of machine learning performance metrics for the validation data, which the model did not see during training. The splitting was stratified such that the training set contained 80% of each species and the validation set contained 20% of each species.

4.5 Training and evaluation of the machine learning models

In this section, the two main development phases for the project are described: the preliminary modelling phase and the secondary modelling phase. Random Forest (Breiman 2001) was chosen as the machine learning model architecture and the overall accuracy was chosen as the model performance metric. The results of the tests and the final model are presented in Section 5. These tests allow us to examine the main factors influencing the accuracy of the predictions and design a final approach that achieves a high overall accuracy using many carefully selected training samples and a spatially consistent thematic map.

During the preliminary modelling phase, the project team focused on refining and testing the pixel-based approach, as it produced higher accuracy scores than the object-based approach when dealing with a small sample size. During the secondary modelling phase, the project team developed a method to increase the number of samples, and consequently shifted focus to the object-based approach, which produced a more spatially consistent thematic map.

4.5.1 Preliminary modelling phase

The goal of the preliminary modelling phase was to investigate how well species could be classified with existing samples, using only the remote sensing and ancillary data features for the classification year (2022). The results would then inform the design of the final model. To this end, five model configurations were tested, which are described below.

Baseline models for object-based and pixel-based approaches

As a baseline, both an object-based and a pixel-based method were trained. The object-based method used the segmentation approach described in Section 4.4.1, which naturally included the ability to calculate spatial statistics features for each object – referred to as the object-based model. Equivalent features were generated in the pixel-based approach by taking spatial statistics for pixels within a radius of 50m of the pixel being classified – referred to as the pixel-based model. Both models used plots that had confidence scores of 3, 4, or 5. Both models used environmental and spectral features, after removing features with a correlation score of 0.9 or above. These models were used to assess the spatial properties of the species extent maps produced by the pixel-based and object-based methods, as well as the overall accuracy generated when using both methods on the same set of samples.

Model with key species and a grouped off-target class

The baseline pixel-based model was altered to explicitly predict the three species that are required for reporting against outcomes of the 2019 BWS (*Eucalyptus camaldulensis*, *Eucalyptus coolabah*, *Eucalyptus largiflorens*, referred hereafter to as the three key species) along with a single off-target “other” class that contained all other selected species – referred to as the pixel-based model with four classes. This model was used to assess the accuracy that could be achieved when grouping all species other than the three key species, compared to modelling all species explicitly.

Model with high confidence scores

The baseline pixel-based model was altered to only use plots with a confidence score of 5 - referred to as the pixel-based model with only confidence scores of 5. This model was used to assess the accuracy that could be achieved when only using high-confidence samples, compared to the collection of mid- to high-confidence samples.

Model with only spectral features

The baseline pixel-based model was altered to only use spectral features, removing all environmental features – referred to as the pixel-based model without environmental features. This model was used to assess how spectral features influenced the spatial properties of the species extent maps.

4.5.2 Secondary modelling phase

The goal of the secondary modelling phase was to improve the model using low-cost approaches including:

- Maximising the number of training samples for machine learning.
- Removing the model features and species that had inconsistent statistical distributions across 2019 to 2022.

With a higher number of samples available, the project team chose to use the object-based approach during this phase as this mitigated the challenges Cunningham et al. (2013) experienced with scattered areas of *Eucalyptus coolabah* in the southern Basin. To model configurations were tested, as described below.

Model with extended sampling

In this model, species samples from all New South Wales and Victoria were incorporated, keeping samples from South Australia and Queensland restricted to the Basin. Each year of Earth observation data was then treated as an independent observation of the samples. It was assumed that the recorded species was present during all sample years; for example, if a field collection team logged the presence of *Casuarina cunninghamiana* in 2020, it was assumed that it was also present at that location in 2019, 2021, and 2022.

Model with extended sampling, removal of summer features, and reduced species

After increasing the sample size, the statistical distributions of key features for each species across each year of the sampling were reviewed. The goal was to remove model features and species that showed high variance year-on-year. The project team hypothesised that removing these sources of noise would make the model more robust to annual changes in weather patterns. For this model, all summer features and four species were removed (discussed in Section 4.4.2, Feature engineering: Step 4 and Step 5).

4.5.3 Final model

A TensorFlow RF model was constructed using 80% of the final sample data for model training, while the remaining 20% was used for model validation. The project team used 2,000 decision trees in the model to ensure high precision in the class likelihood estimates. The number of attributes to assess at each split was set to 20 using the established convention of the square root of all available features (434). A “best first global” approach was applied for tree construction over a “local” approach to target node splits that provided the greatest overall model improvement. The maximum depth of any given tree was set to 16 with no constraint on the number of nodes. These settings are summarised in Table 3.

Table 3. TensorFlow random forest hyper-parameter settings and values for the final model.

Hyper-Parameter	Value
num_trees	2,000
growing_strategy	BEST_FIRST_GLOBAL
max_depth	16
max_num_nodes	-1
num_candidate_attributes	20
split_axis	AXIS_ALIGNED

The output of the model is a likelihood layer for each species. For each pixel, the likelihood layer for a given species captures the proportion of decision trees in the random forest that classified the pixel as that species. The value of the layer ranges from zero to one, where zero indicates that no random forest trees classified the pixel as the given species, and one indicating that all trees classified the pixel as the given species.

4.6 Generation of the Basin-wide map

The Basin-wide map for 2022 was generated using a combination of Google Earth Engine to organise data and VertexAI to execute the classification in parallel. Google Earth Engine was used to generate the inputs for the classification model, including layers from image segmentation, temporal aggregations, and the spatial statistics for each object. The layers were created as-needed and retained in memory without the construction of permanent assets. VertexAI was used to deploy the pretrained final model at endpoints accessible to the Google Earth Engine infrastructure. A total of four VertexAI model endpoints were deployed to support parallel processing of the classification task.

The large scale and volume of the data to be classified required that data be served from Google Earth Engine to the VertexAI model end point using tiling. A tile size of 20x20km was selected, producing 3,333 contiguous tiles covering the Basin, as shown in Figure 8. Tiled data were consecutively served to one of the four available model endpoints for image classification. The resulting likelihood layers were exported back to Google Earth Engine as an Image Asset within an Image Collection for post-processing.

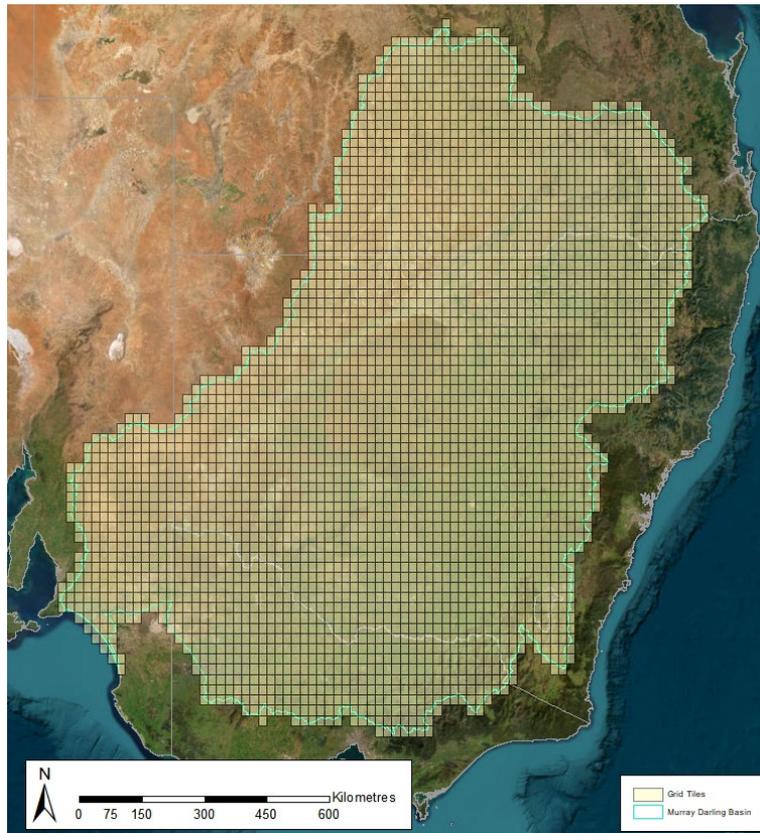


Figure 8. Tiling system used to process the entire Murray-Darling Basin, where each tile is a 20x20km square. The outline of the Basin is shown in green.

4.7 Post processing

In the post-processing, Dynamic World serves two purposes. Firstly, it is used to identify off-target landcover classes, and secondly, it is used to collapse low confidence hard classifications to a more broader vegetation group. Using the layers generated in the previous step, a thematic classified map was produced by performing a hard classification of the species likelihood layers along with the Dynamic World land cover dataset. Dynamic World was first used to mask irrelevant land cover classes, excluding areas where the species probability model was unlikely to give meaningful predictions. This masking strategy kept all predictions that overlapped with a Dynamic World on-target class (flooded vegetation, shrub and scrub, and trees), and masked all predictions that overlapped with a Dynamic World off-target class (bare, built, crops, grass, and water). Finally, objects with low confidence hard classifications predictions from the model were assigned to a broader class, such as Trees, rather than specific species.

4.7.1 Hard classification

A hard classification was produced by collapsing the likelihood layers into a single layer. For all pixels in each object, the per-species likelihood measures were extracted and compared. The pixels in each object were assigned to the species with the highest likelihood measure for that object. This was later post-processed to both remove irrelevant land cover, and to identify areas of low model likelihood that could require additional post-processing. The correspondence of species to pixel value in the hard classification map is given in Table 5.

4.7.2 On-target and off-target land cover identification

Dynamic World was used to identify on-target and off-target land cover. The Dynamic World dataset is a time-series of Sentinel-2 classifications. The nine classes of this classification were remapped to two classes: on-target and off-target. When reviewing Dynamic World across the Murray–Darling Basin, the project team found that the “Grass” class primarily captured pasture land as opposed to native grassland. As such, “Grass” was assigned to the off-target class. The allocation of Dynamic World classes to on- and off-target are given in Table 4.

Table 4. Assignment of Dynamic World classes to on-target and off-target classes.

On-Target	Off-Target
Flooded Vegetation	Bare
Shrub and Scrub	Built
Trees	Crops
	Grass
	Water

To create an annual on- and off-target layer, the project team collated the 2022 Dynamic World classifications across the study site and remapped their class values. The frequency of classifications that corresponded to any of the on-target Dynamic World classes was calculated. On-target objects were defined as those with an on-target frequency of at least 80% across the year. Areas that did not meet this requirement were classified as off-target. Off-target areas were masked out of the hard classification, producing an extent map.

4.7.3 Broader class collapse

Where the classification model developed for this work focuses on specific vegetation species, Dynamic World favours broader, more generalised vegetation groups. To produce a final species extent map, Dynamic World was used to remap objects with low model confidence in the hard classification map (Section 4.7.1) to a broader vegetation group. It was found that a likelihood confidence threshold of 30% was sufficient for identifying confused or incorrectly classified pixels. Pixels with values that fell beneath this threshold were defined as low confidence pixels. These pixels were then remapped to their broader Dynamic World counterpart class. The final pixel values for the classes are given in Table 5.

Table 5. Mapping of species to pixel value in the post-processed classification map.

Class	Pixel Value	Class	Pixel Value
<i>Acacia salicina</i>	1	<i>Eucalyptus largiflorens</i>	10
<i>Atriplex semibaccata</i>	2	<i>Eucalyptus melliodora</i>	11
<i>Callitris glaucophylla</i>	3	<i>Eucalyptus microcarpa</i>	12
<i>Casuarina cunninghamiana</i>	4	<i>Lycium ferocissimum</i>	13
<i>Duma florulenta</i>	5	<i>Nitraria billardierei</i>	14
<i>Eucalyptus albens</i>	6	(DW) Trees	15
<i>Eucalyptus blakelyi</i>	7	(DW) Flooded Vegetation	16
<i>Eucalyptus camaldulensis</i>	8	(DW) Shrub and Scrub	17
<i>Eucalyptus coolabah</i>	9		

4.8 Change detection

4.8.1 Overview

Change detection is valuable for measuring how the extent of species across the Basin have been affected by specific events, such as fire and deforestation, but also for monitoring recruitment over time and the impacts of MDBA policies. Change detection using thematic maps between two dates is straight-forward, a pixel either has changed label or has not changed label. In this case however, there is no deeper intuition into the magnitude of change, with no distinction between subtle shifts and blunt changes, or simply model confusion.

Instead of the simplistic approach above, a method was developed that accounts for the properties of the species extent map, namely that the key product is the species likelihood layers. While this method is more nuanced than a comparison of two hard-classification maps, it can produce insights into the magnitude of change and the level of confidence that the change has really occurred.

There are a broad range of methods to approach change detection. For this work, one approach is illustrated, but the project team note that the MDBA may need to customise this based on the specific change detection application. To illustrate change detection required a second classification layer. A classification of 2021 data was therefore performed over only a few targeted areas. One area is used to illustrate the proof of concept for change detection below.

4.8.2 Change detection approach

The change detection approach for this work has two steps. The first identifies areas of labelled change. The second step explores the confidence surrounding those changes and draws conclusions on the likelihood of change being true change or the result of model confusion. The approach required probability and hard classification layers for both 2021 and 2022. An area of interest was therefore selected and a classification of 2021 data across its extents was performed.

The first step is to identify and define differences between the 2021 and 2022 datasets on a thematic level as either a Change or No Change. Change itself can be further separated into a straight change between two species of interest, a conversion of an off-class land cover to a class of interest, or a conversion of a class of interest to an off-class land cover. No Change occurs where the thematic label is consistent between datasets. The two hard classification datasets are used. As these represent nominal values, a per-pixel binary operation of non-equality is applied. Pixels that are labelled the same will return false, while pixels with different labels will return true. Pixels labelled as true indicate areas that have undergone one of the three forms of change.

The second step is to quantify the model confidence for each changed pixel, which is performed by looking at the difference between the model likelihood values for the classified species in the first year, and the classified species in the second year. If the difference in model likelihood values for the two years is small, there is a higher probability that the detected change is by chance alone. The confidence for such a change may be considered low, and may reflect subtle, insignificant changes or a model misclassification. Conversely, where the difference in model likelihood values for the two years is large, there is a lower probability that the detected change is by chance alone. The confidence for such a change may be considered high and may reflect significant changes in the landscape.

Using the principles above, the ratio of the likelihood values for both species can also be considered. This is done for both years, as illustrated in the figures below. For an area classified as class A in 2021, but classified as class B in 2022, two confusion indices are calculated: the ratio of the 2021 likelihood values for species A and B (Figure 9) and the ratio of the 2022 likelihood values for species A and B (Figure 10). The two indices can then be used to assess the level of confidence of the model prediction in each year; a concrete example is provided in Section 5.4.

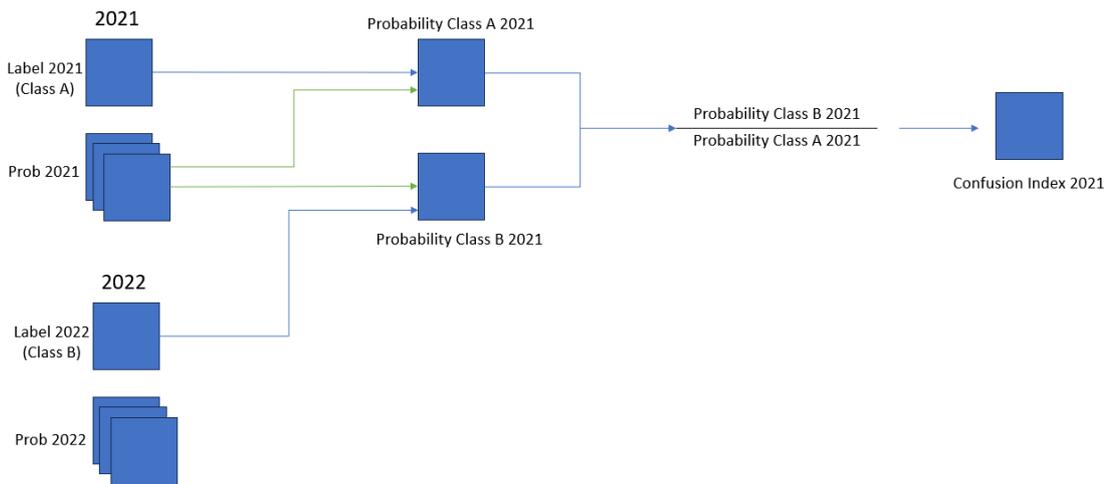


Figure 9. After identifying the pixels that have changed from class A in 2021 to class B in 2022, the confusion index is calculated as the ratio of the class A and B probabilities in 2021. If the model produced similar probabilities for both classes in 2021, the confusion index will be high, and the original classification could be considered low-confidence.

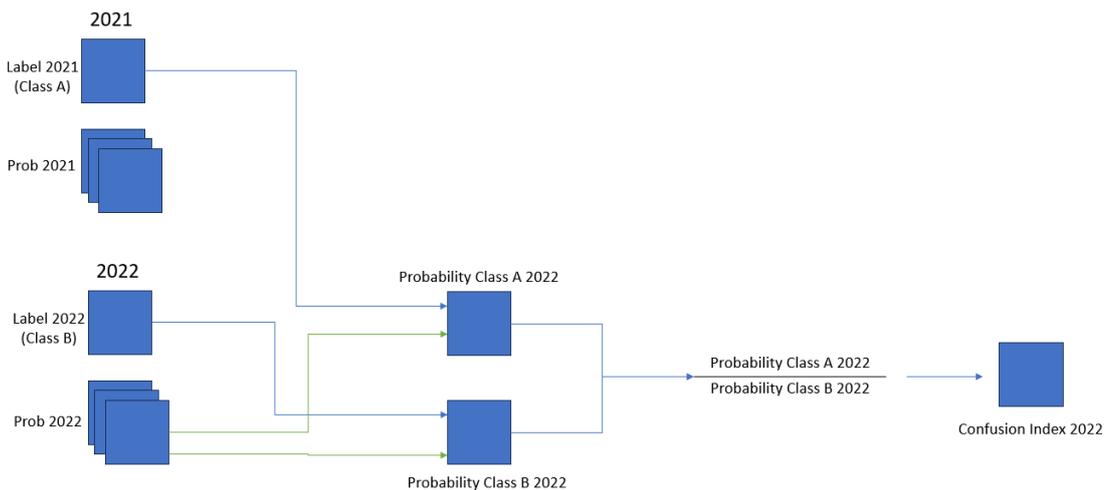


Figure 10. After identifying the pixels that have changed from class A in 2021 to class B in 2022, the confusion index is calculated as the ratio of the class A and B probabilities in 2022. If the model produced similar probabilities for both classes in 2022, the confusion index will be high, and the new classification could be considered low-confidence.

5 Results

In this section, the results from the preliminary, secondary, and final model phases are presented, including the overall accuracy for the tested models. In the case of the final model, the confusion matrix is also presented, along with the commission and omission errors for each species. In all cases, results were produced for a single 20% validation set, which had been withheld from model training.

Following the model performance results, relevant results from the post-processing stage, the final likelihood and hard-classification maps, and the change detection demonstration are presented.

5.1 Model performance scores on validation set

5.1.1 Preliminary models

Table 6 shows the key characteristics and overall accuracy for the five models tested during the preliminary phase (Section 4.5.1). The models have the following characteristics in common:

- All samples were within the Murray–Darling Basin boundary
- For all samples, temporal data from 2022 was used
- When selecting features, a correlation score threshold of 0.9 was used to remove features that were more than 90% correlated with an existing selected feature

Preliminary baseline classes (13): *Acacia stenophylla*, *Atriplex semibaccata*, *Atriplex suberecta*, *Callitris glaucophylla*, *Duma florulenta*, *Eucalyptus albens*, *Eucalyptus blakelyi*, *Eucalyptus camaldulensis*, *Eucalyptus coolabah*, *Eucalyptus largiflorens*, *Eucalyptus melliodora*, *Eucalyptus microcarpa*, *Lycium ferocissimum*

Preliminary reduced classes (4): *Eucalyptus camaldulensis*, *Eucalyptus coolabah*, *Eucalyptus largiflorens*, other

Table 6. Characteristics and overall accuracy for preliminary models

	Object-based	Pixel-based	Pixel-based with four classes	Pixel-based with only confidence scores of 5	Pixel-based without environmental features
Number of classes	13	13	4	13	13
Classes	Baseline	Baseline	Reduced	Baseline	Baseline
Spatial aggregation	Object-based	Pixel-based	Pixel-based	Pixel-based	Pixel-based
Sample confidence	3, 4, 5	3, 4, 5	3, 4, 5	5	3, 4, 5
Number of samples	6,607	9,064	9,064	8,004	9,064
Type of features	Spectral and environmental	Spectral and environmental	Spectral and environmental	Spectral and environmental	Spectral only
Overall accuracy	63.4%	75.1%	88.8%	77.9%	68.1%

5.1.2 Secondary models

Table 7 shows the key characteristics and overall accuracy for the two models tested during the secondary phase (Section 4.5.2), along with the preliminary object-based model for comparison. Note that while the primary focus was on the pixel-based model in the preliminary phase, focus was switched to the object-based model in the secondary phase once a larger training sample was available.

Preliminary baseline classes (13): *Acacia stenophylla*, *Atriplex semibaccata*, *Atriplex suberecta*, *Callitris glaucophylla*, *Duma florulenta*, *Eucalyptus albens*, *Eucalyptus blakelyi*, *Eucalyptus camaldulensis*, *Eucalyptus coolabah*, *Eucalyptus largiflorens*, *Eucalyptus melliodora*, *Eucalyptus microcarpa*, *Lycium ferocissimum*

Secondary baseline classes (18): *Acacia salicina*, *Acacia stenophylla*, *Atriplex nummularia*, *Atriplex semibaccata*, *Atriplex suberecta*, *Callitris glaucophylla*, *Casuarina cunninghamiana*, *Chenopodium nitrariaceum*, *Duma florulenta*, *Eucalyptus albens*, *Eucalyptus blakelyi*, *Eucalyptus camaldulensis*, *Eucalyptus coolabah*, *Eucalyptus largiflorens*, *Eucalyptus melliodora*, *Eucalyptus microcarpa*, *Lycium ferocissimum*, *Nitraria billardierei*

Secondary reduced classes (14): *Acacia salicina*, *Atriplex semibaccata*, *Callitris glaucophylla*, *Casuarina cunninghamiana*, *Duma florulenta*, *Eucalyptus albens*, *Eucalyptus blakelyi*, *Eucalyptus camaldulensis*, *Eucalyptus coolabah*, *Eucalyptus largiflorens*, *Eucalyptus melliodora*, *Eucalyptus microcarpa*, *Lycium ferocissimum*, *Nitraria billardierei*

Table 7. Characteristics and overall accuracy for secondary models

	Preliminary object-based	Object-based with extended sampling	Object-based with extended sampling, no summer features, and reduced species
Number of classes	13	18	14
Classes	Preliminary baseline	Secondary baseline	Secondary reduced
Spatial aggregation	Object-based	Object-based	Object-based
Sample confidence	3, 4, 5	5	5
Temporal sampling	2022	2019-2022	2019-2022
Sample area	Murray-Darling Basin	MDB + all of Vic and NSW	MDB + all of Vic and NSW
Number of samples	6,607	19,806	19,110
Correlation threshold	0.9	0.99	0.99
Number of features	73	562	434
Type of features	Spectral and environmental	Spectral and environmental	Spectral and environmental, no summer features
Overall accuracy	63.4%	77.0%	80.3%

5.1.3 Final model

The final model chosen was the object-based model with extended sampling, 14 species, and no summer variables (Section 4.5.3). Validating this model on 20% of the 19,110 samples gave an overall accuracy of 80.3% and produced the following confusion matrix (Table 8) and omission and commission errors (Table 9).

Table 8. Confusion matrix for the final model, evaluated against 3,822 samples from 14 species.

		Truth													
		<i>Acacia salicina</i>	<i>Atriplex semibaccata</i>	<i>Callitris glaucophylla</i>	<i>Casuarina cunninghamiana</i>	<i>Duma florulenta</i>	<i>Eucalyptus albens</i>	<i>Eucalyptus blakelyi</i>	<i>Eucalyptus camaldulensis</i>	<i>Eucalyptus coolabah</i>	<i>Eucalyptus largiflorens</i>	<i>Eucalyptus melliodora</i>	<i>Eucalyptus microcarpa</i>	<i>Lycium ferocissimum</i>	<i>Nitraria billardierei</i>
Predicted	<i>Acacia salicina</i>	44	1	0	2	0	0	0	0	0	0	0	0	0	0
	<i>Atriplex semibaccata</i>	0	93	3	1	5	0	0	0	0	2	2	0	5	0
	<i>Callitris glaucophylla</i>	2	16	603	1	6	37	13	10	5	4	8	22	26	1
	<i>Casuarina cunninghamiana</i>	0	4	0	186	0	0	1	2	0	0	4	4	5	0
	<i>Duma florulenta</i>	0	1	0	0	56	0	0	2	1	6	0	0	0	0
	<i>Eucalyptus albens</i>	3	2	24	7	0	363	29	5	0	0	38	5	2	0
	<i>Eucalyptus blakelyi</i>	1	1	9	7	0	16	274	0	0	0	37	6	4	0
	<i>Eucalyptus camaldulensis</i>	1	8	0	4	7	7	8	247	0	1	16	14	10	0
	<i>Eucalyptus coolabah</i>	3	0	1	0	8	0	0	3	77	2	0	0	3	0
	<i>Eucalyptus largiflorens</i>	0	1	2	0	12	0	0	7	1	106	0	0	0	2
	<i>Eucalyptus melliodora</i>	0	10	2	8	0	6	32	4	0	0	380	6	7	0
	<i>Eucalyptus microcarpa</i>	0	3	5	0	0	5	0	4	0	1	6	128	4	0
	<i>Lycium ferocissimum</i>	13	41	2	14	9	3	9	5	6	13	5	2	453	0
	<i>Nitraria billardierei</i>	0	0	2	0	0	0	0	0	0	3	0	0	3	60

Table 9. The omission and commission rates from the final model for all 14 species. The omission error measures the fraction of samples that truly belong to a class but were predicted as a different class. The commission error measures the fraction of samples that were predicted as a class, but do not truly belong to that class.

Species	Number of true samples	Omission error	Commission error	Species	Number of true samples	Omission error	Commission error
<i>Acacia salicina</i>	67	34.3%	6.4%	<i>Eucalyptus camaldulensis</i>	289	14.5%	23.5%
<i>Atriplex semibaccata</i>	181	48.6%	16.2%	<i>Eucalyptus coolabah</i>	90	14.4%	20.6%
<i>Callitris glaucophylla</i>	653	7.7%	20.0%	<i>Eucalyptus largiflorens</i>	138	23.2%	19.1%
<i>Casuarina cunninghamiana</i>	230	19.1%	9.7%	<i>Eucalyptus melliodora</i>	496	23.4%	16.5%
<i>Duma florulenta</i>	103	45.6%	15.2%	<i>Eucalyptus microcarpa</i>	187	31.6%	17.9%
<i>Eucalyptus albens</i>	437	16.9%	24.1%	<i>Lycium ferocissimum</i>	522	13.2%	21.2%
<i>Eucalyptus blakelyi</i>	366	25.1%	22.8%	<i>Nitraria billardierei</i>	63	4.8%	11.8%

Table 10. Comparison of omission and commission errors on species in common between this work and Cunningham et al. (2013).

Species	Omission Error		Commission Error	
	Cunningham et al. (2013)	This work	Cunningham et al. (2013)	This work
<i>Casuarina cunninghamiana</i>	0.918	0.191	0.328	0.097
<i>Duma florulenta</i>	0.378	0.456	0.515	0.152
<i>Eucalyptus camaldulensis</i>	0.163	0.145	0.397	0.235
<i>Eucalyptus coolabah</i>	0.110	0.144	0.568	0.206
<i>Eucalyptus largiflorens</i>	0.228	0.232	0.734	0.191

Feature importance

The 20 most important features by rank are presented in Table 11, with the 20 most important biophysical features by rank in Table 12. A visual representation of the rank of all 434 features by type is presented in Figure 11.

Table 11. The top 20 features (of any kind) as ranked by importance score in the final model. The spatial statistic calculated for the object is given in brackets.

Rank	Feature	Rank	Feature
1	SRTM Elevation (Maximum)	11	SLGA Soil Depth (Mean)
2	Winter, SILO Max Temperature (Maximum)	12	Autumn, SILO Evaporation (Maximum)
3	Winter, SILO Min Temperature (Maximum)	13	SLGA Soil Depth (Minimum)
4	SLGA Soil Carbon Content (Mean)	14	SLGA Soil Silt Content (Maximum)
5	SLGA Soil Carbon Content (Minimum)	15	SLGA Soil Clay Content (Maximum)
6	SLGA Soil Carbon Content (Maximum)	16	SLGA Soil Clay Content (90th Percentile)
7	Spring, SILO Average Rainfall (Maximum)	17	SLGA Soil Depth (Maximum)
8	Autumn, SILO Min Temperature (Maximum)	18	SLGA Soil Sand Content (Minimum)
9	SLGA Soil Silt Content (Mean)	19	SRTM Slope (Maximum)
10	SLGA Soil Silt Content (Minimum)	20	SLGA Soil Clay Content (Mean)

Table 12. The top 20 biophysical features as ranked by importance score in the final model. The spatial statistic calculated for the object is given in brackets.

Rank	Feature	Rank	Feature
41	Winter, Sentinel-2 NIR (Standard Deviation)	69	Autumn, Sentinel-2 Blue (Maximum)
46	Spring, Sentinel-2 NDMI (Maximum)	70	Winter, Sentinel-2 Green (Minimum)
52	Autumn, Sentinel-2 NIR (Standard Deviation)	73	Autumn, Sentinel-2 NDMI (Standard Deviation)
57	Autumn, Sentinel-2 Blue (Standard Deviation)	75	Autumn, Sentinel-2 NDWI (Maximum)
58	Autumn, Sentinel-2 Blue (90th Percentile)	77	Spring, Sentinel-2 SWIR 2 (Minimum)
60	Spring, Sentinel-2 NDMI (Standard Deviation)	78	Spring, Sentinel-2 NDWI (Minimum)
61	Spring, Sentinel-2 NDWI (Standard Deviation)	79	Winter, Sentinel-2 SWIR 1 (Minimum)
62	Winter, Sentinel-2 NDMI (Standard Deviation)	80	Winter, Sentinel-2 Vegetation Red Edge 2 (Standard Deviation)
64	Autumn, Sentinel-2 NDWI (Standard Deviation)	81	Winter, Sentinel-2 NDMI (Maximum)
67	Winter, Sentinel-2 SWIR 2 (Minimum)	83	Spring, Sentinel-2 NIR (Standard Deviation)

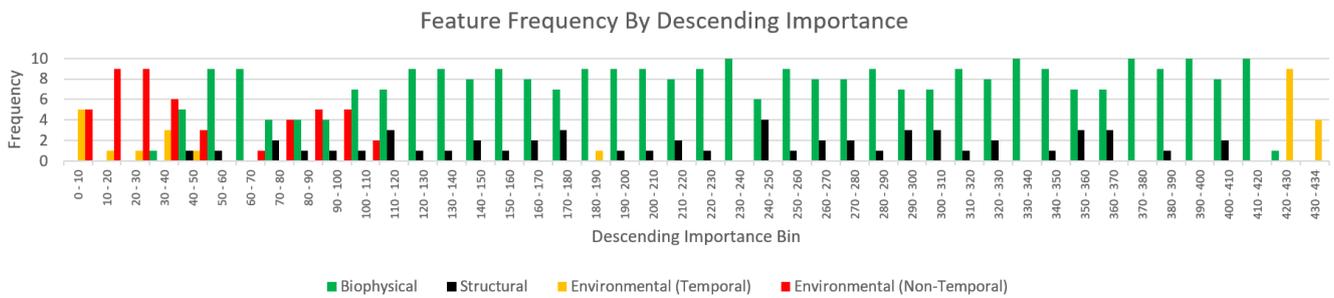


Figure 11. A visual summary showing the number of biophysical, structural, and environmental features by rank, in bins of 10 features. The first bin shows that the model ranked five Environmental (Temporal) and five Environmental (Non-Temporal) features in the top ten features by rank.

5.2 Post processing

To evaluate the suitability of Dynamic World for masking off-target classes, and generalising target classes in cases of low model confidence, the proportion of samples for each species that corresponded to each Dynamic World class was examined, and is shown in Table 13.

Table 13. Proportion of samples that corresponded to each Dynamic World class, grouped by off-target and on-target.

	Off-Target Classes					On-Target Classes			Total Off-Target	Total On-Target
	Bare	Built	Crops	Grass	Water	Flooded Vegetation	Shrub and Scrub	Trees		
<i>Acacia salicina</i>	0.03	0.02	0.1	0.14	0	0	0.17	0.54	0.29	0.71
<i>Atriplex semibaccata</i>	0.03	0.1	0.09	0.13	0.01	0.05	0.25	0.34	0.36	0.64
<i>Callitris glaucophylla</i>	0.04	0.01	0.01	0.05	0	0	0.19	0.7	0.11	0.89
<i>Casuarina cunninghamiana</i>	0.09	0.17	0.02	0.11	0.01	0	0.02	0.58	0.4	0.6
<i>Duma florulenta</i>	0.08	0.01	0.07	0.04	0.12	0.19	0.19	0.3	0.32	0.68
<i>Eucalyptus albens</i>	0.04	0	0.01	0.24	0	0	0.12	0.59	0.29	0.71
<i>Eucalyptus blakelyi</i>	0	0.01	0	0.21	0	0	0.12	0.66	0.22	0.78
<i>Eucalyptus camaldulensis</i>	0.01	0.05	0.03	0.05	0.04	0.04	0.12	0.66	0.18	0.82
<i>Eucalyptus coolabah</i>	0	0	0.02	0.02	0.01	0.06	0.18	0.71	0.05	0.95
<i>Eucalyptus largiflorens</i>	0.11	0	0.01	0	0.03	0.1	0.57	0.18	0.15	0.85
<i>Eucalyptus melliodora</i>	0	0.08	0.05	0.2	0	0	0.13	0.54	0.33	0.67
<i>Eucalyptus microcarpa</i>	0	0.07	0.24	0.08	0	0	0.1	0.51	0.39	0.61
<i>Lycium ferocissimum</i>	0.01	0.13	0.09	0.21	0.01	0.03	0.14	0.38	0.45	0.55
<i>Nitraria billardierei</i>	0.26	0.01	0.15	0.01	0.01	0.04	0.44	0.08	0.44	0.56

5.3 Species mapping

This section includes the Basin-wide likelihood and hard classification maps for 2022, as well as maps for the three key species. These maps have been provided to the MDBA as a deliverable of this project. A collection of local visual inspection sites is also included, which were checked by the MDBA prior to running the final model across the Basin.

5.3.1 Basin-wide mapping

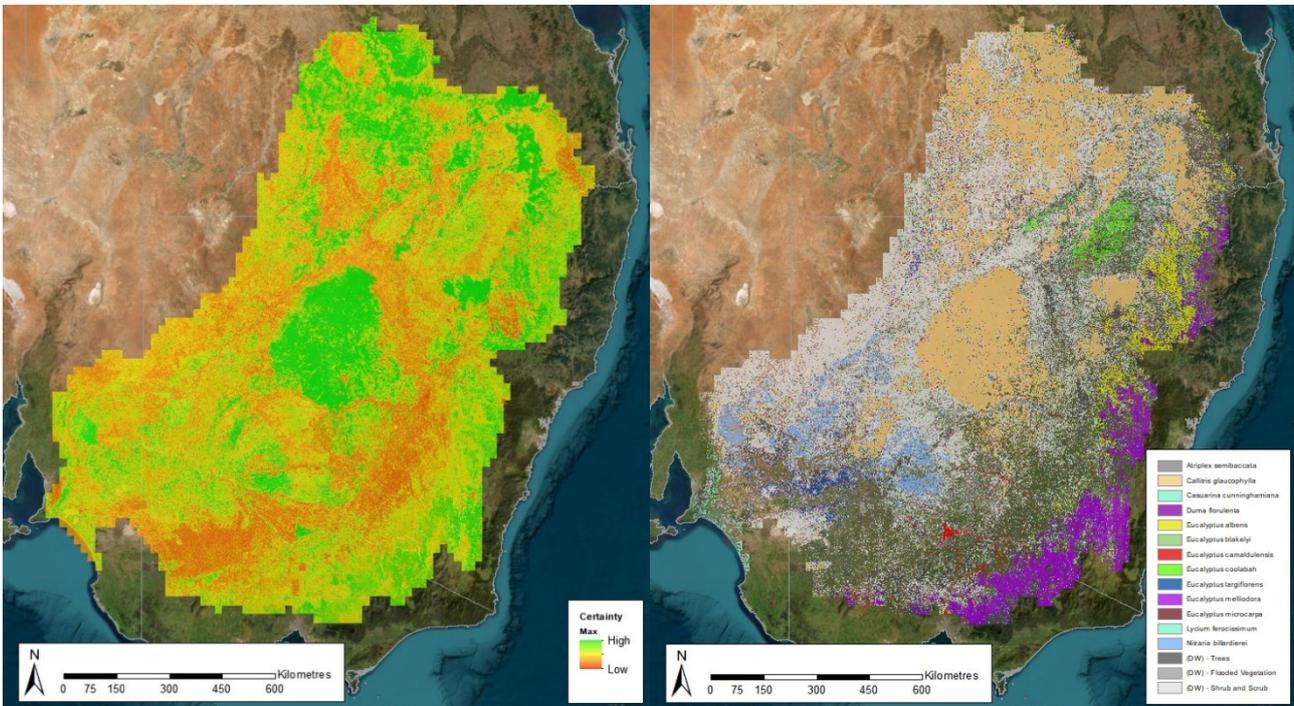


Figure 12. Basin-wide maps. Left: Maximum likelihood value for all pixels in the Basin. Right: Hard classification including masking with Dynamic World, including remapping of areas with less than 30% likelihood to the Dynamic World classification.

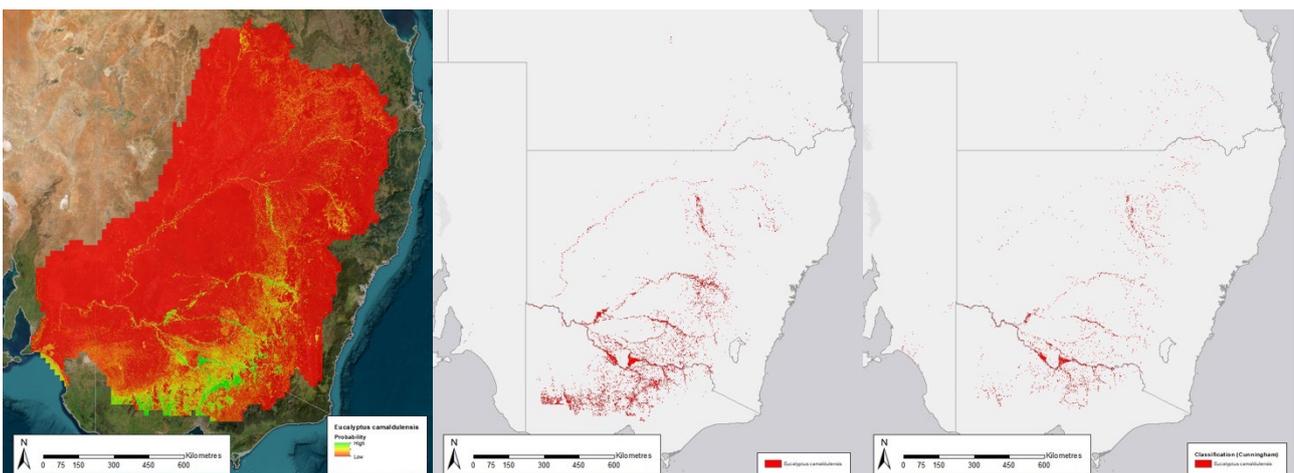


Figure 13. Results for *Eucalyptus camaldulensis*. Left: Likelihood values. Middle: Hard classification. Right: Distribution mapped by Cunningham et al. (2013).

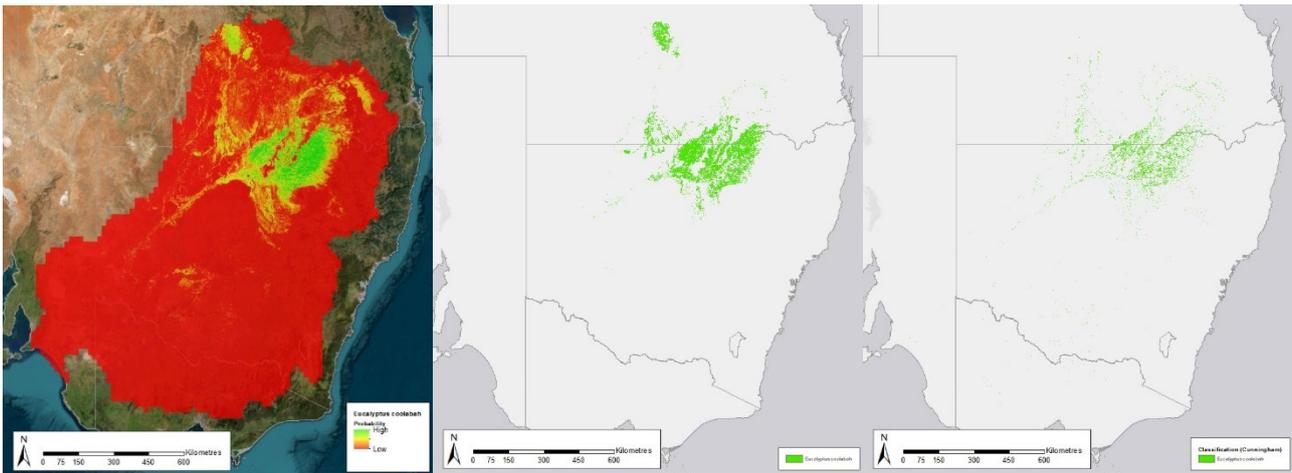


Figure 14. Results for *Eucalyptus coolabah*. Left: Likelihood values. Middle: Hard classification. Right: Distribution mapped by Cunningham et al. (2013).

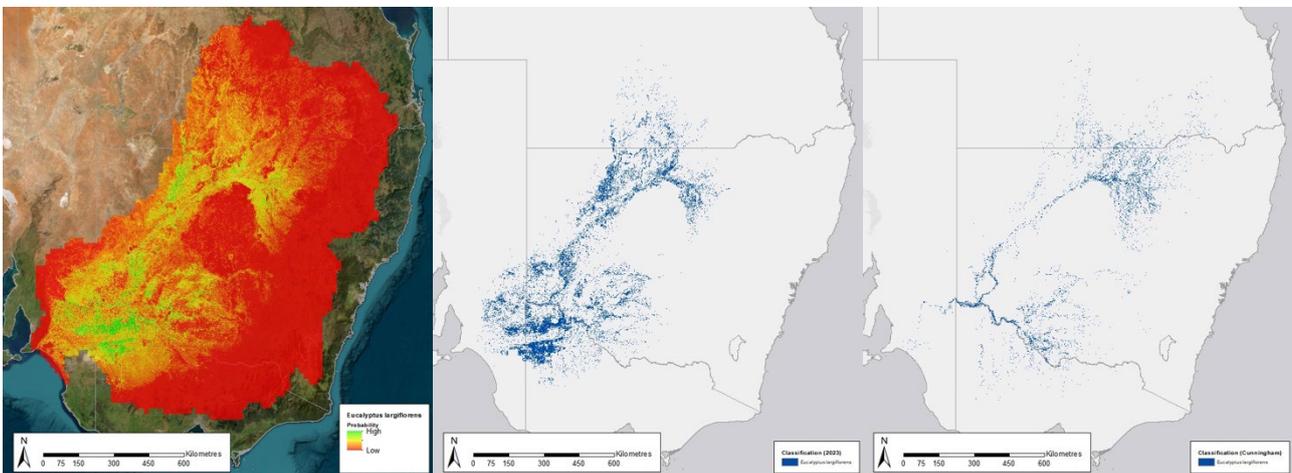


Figure 15. Results for *Eucalyptus largiflorens*. Left: Likelihood values. Middle: Hard classification. Right: Distribution mapped by Cunningham et al. (2013).

5.3.2 Desktop inspection sites

Desktop inspection sites were selected to qualitatively validate the mapping quality of the final model, and were drawn from suggestions from the MDBA, such as sites monitored by the Flow-MER program (e.g. Moree) or areas with well understood species distribution (e.g. Barmah Forest). The locations of the selected sites are shown in Figure 16, followed by the 2022 hard classification for each site.

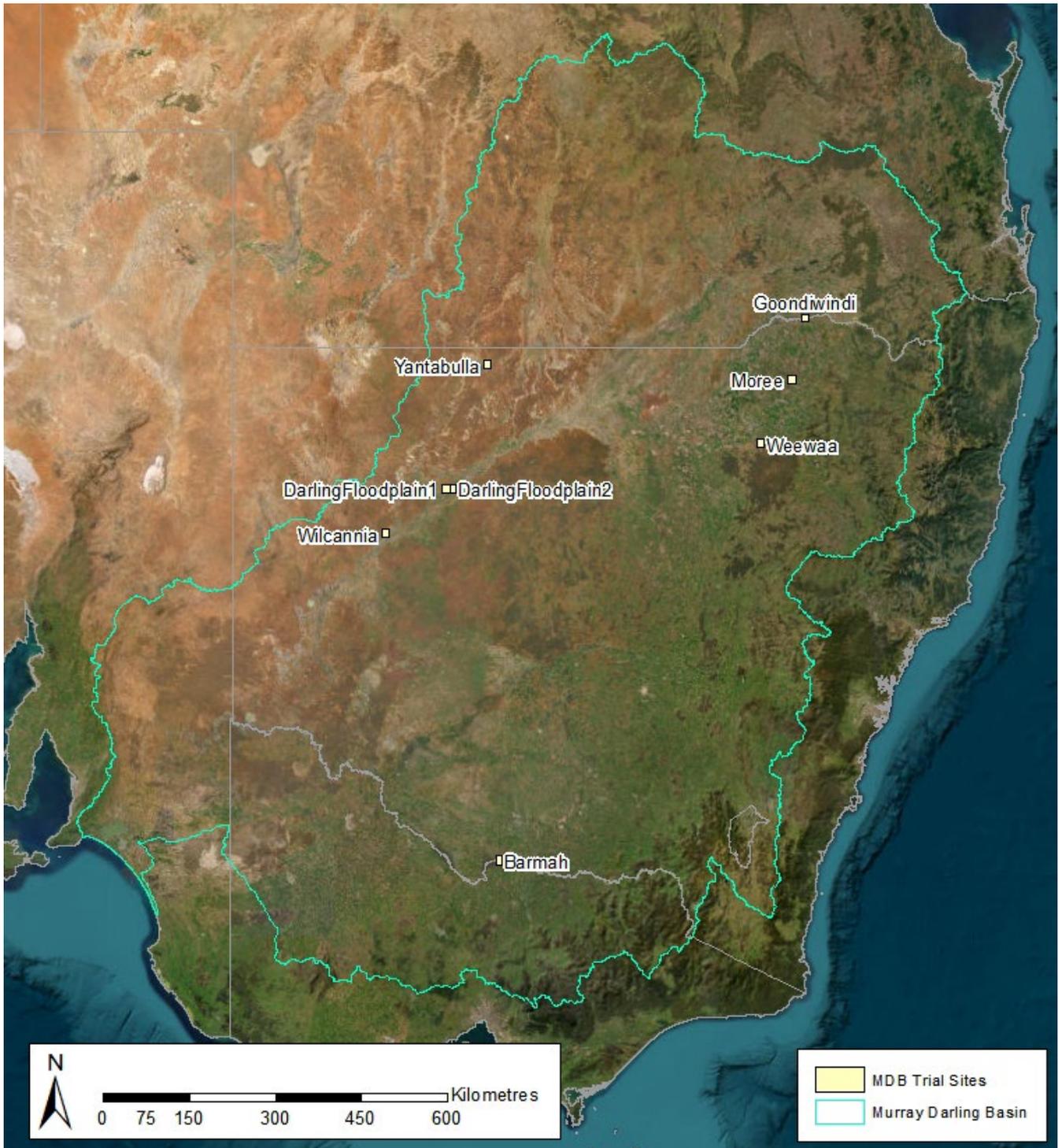


Figure 16. Selected desktop inspection sites for the 2022 hard classification map.

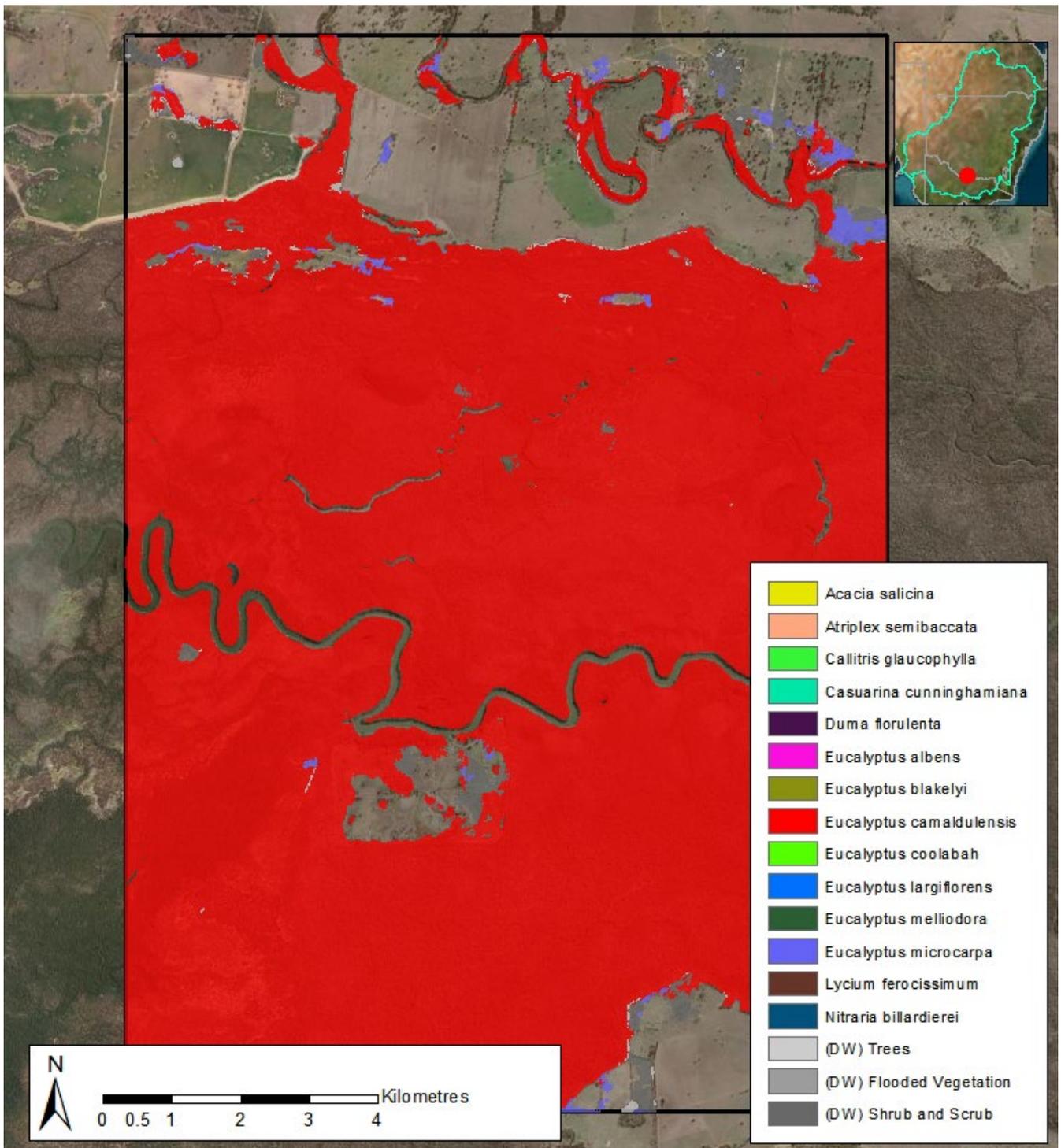


Figure 17. Desktop inspection site 1: Barmah Forest

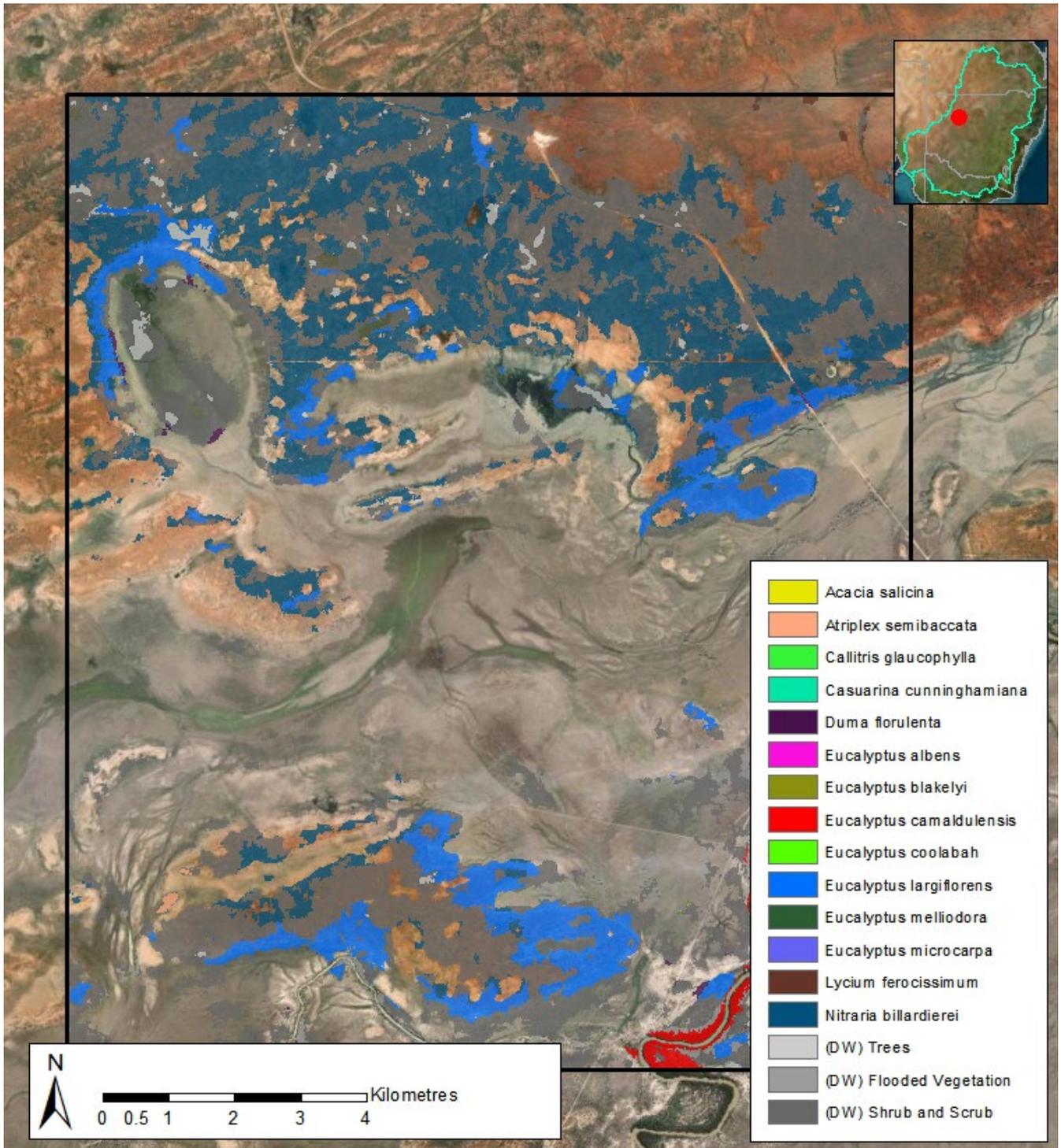


Figure 18. Desktop inspection site 2: Darling Floodplain 1

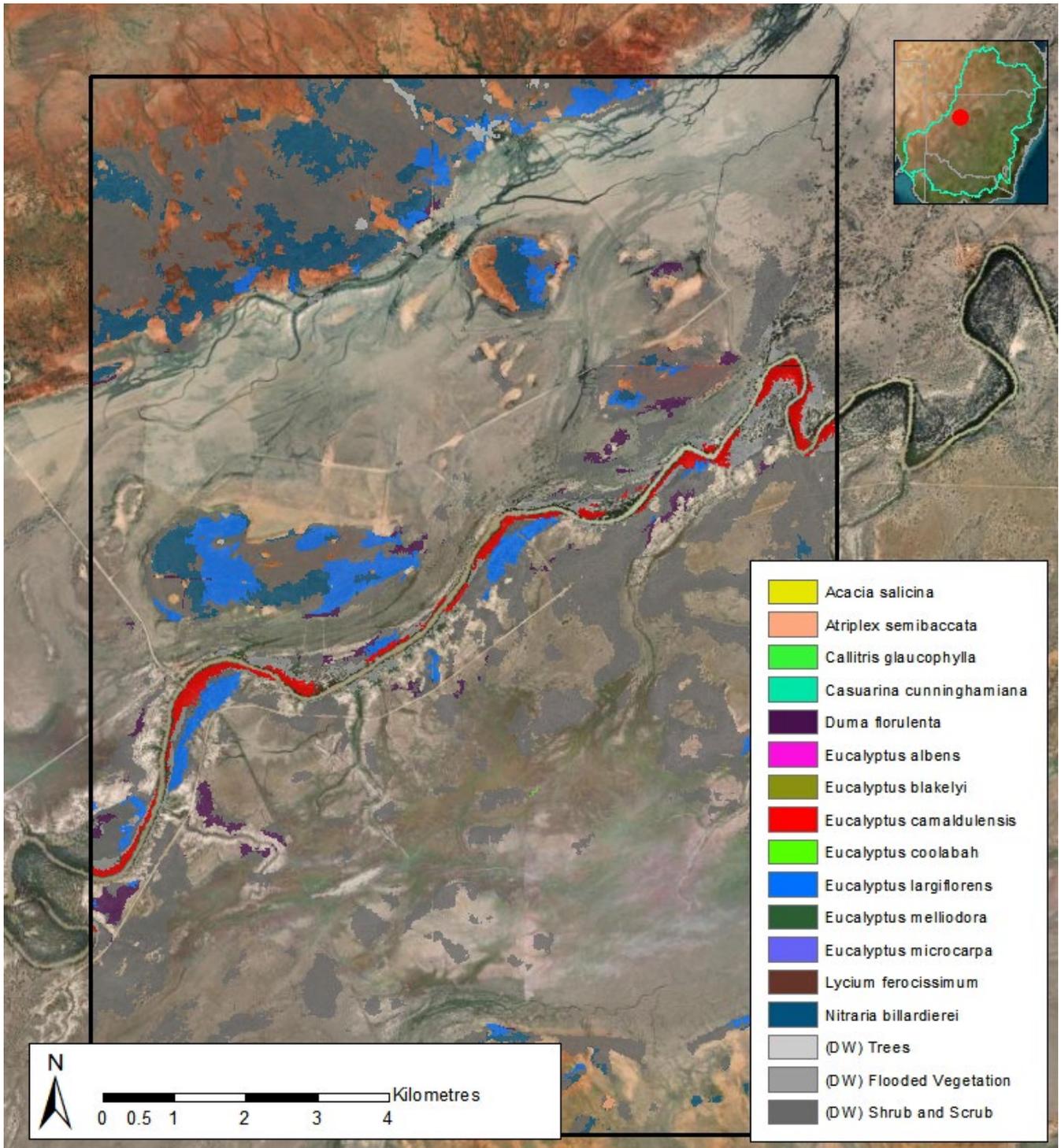


Figure 19. Desktop inspection site 3: Darling Floodplain 2

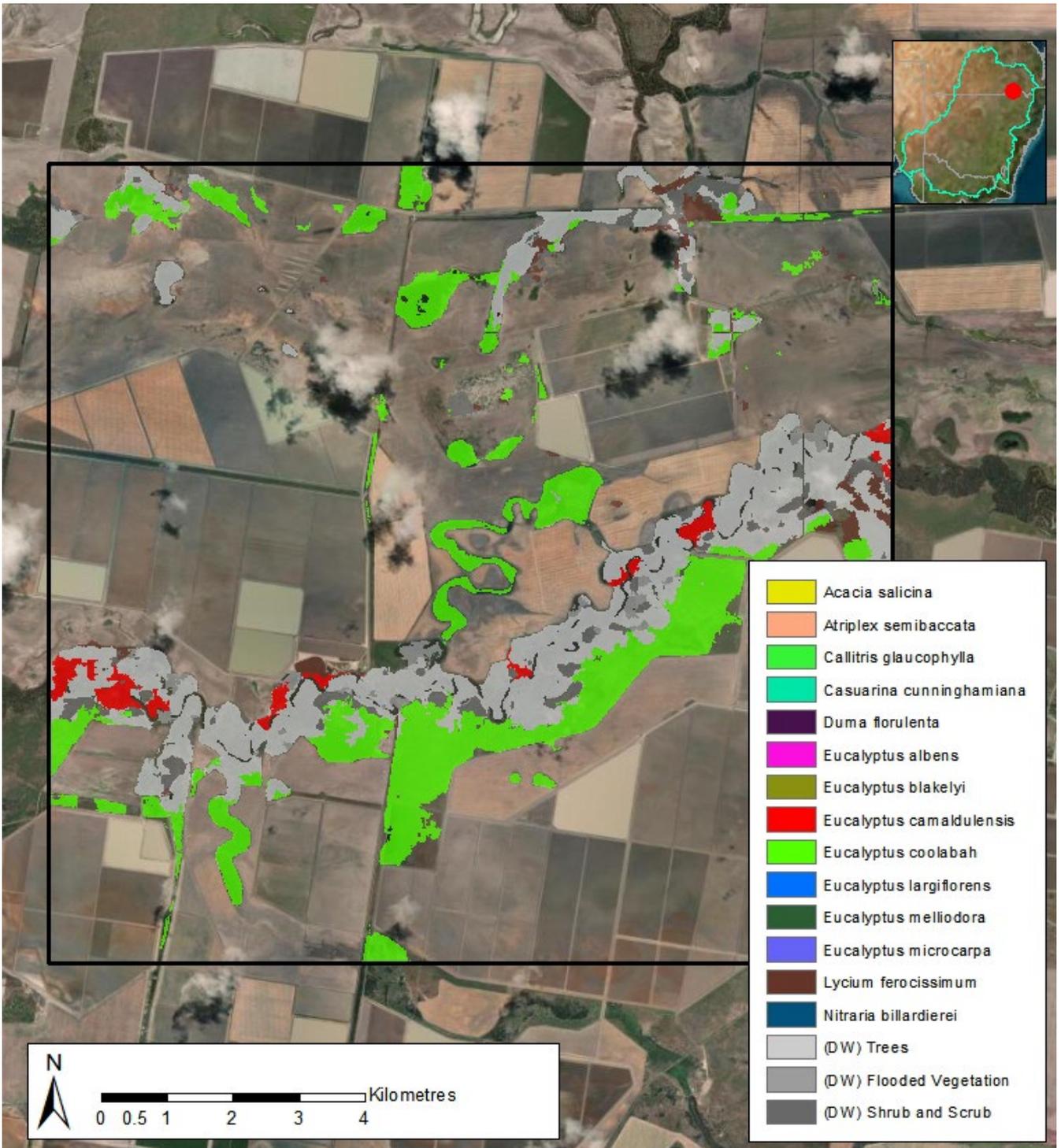


Figure 20. Desktop inspection site 4: Goondiwindi

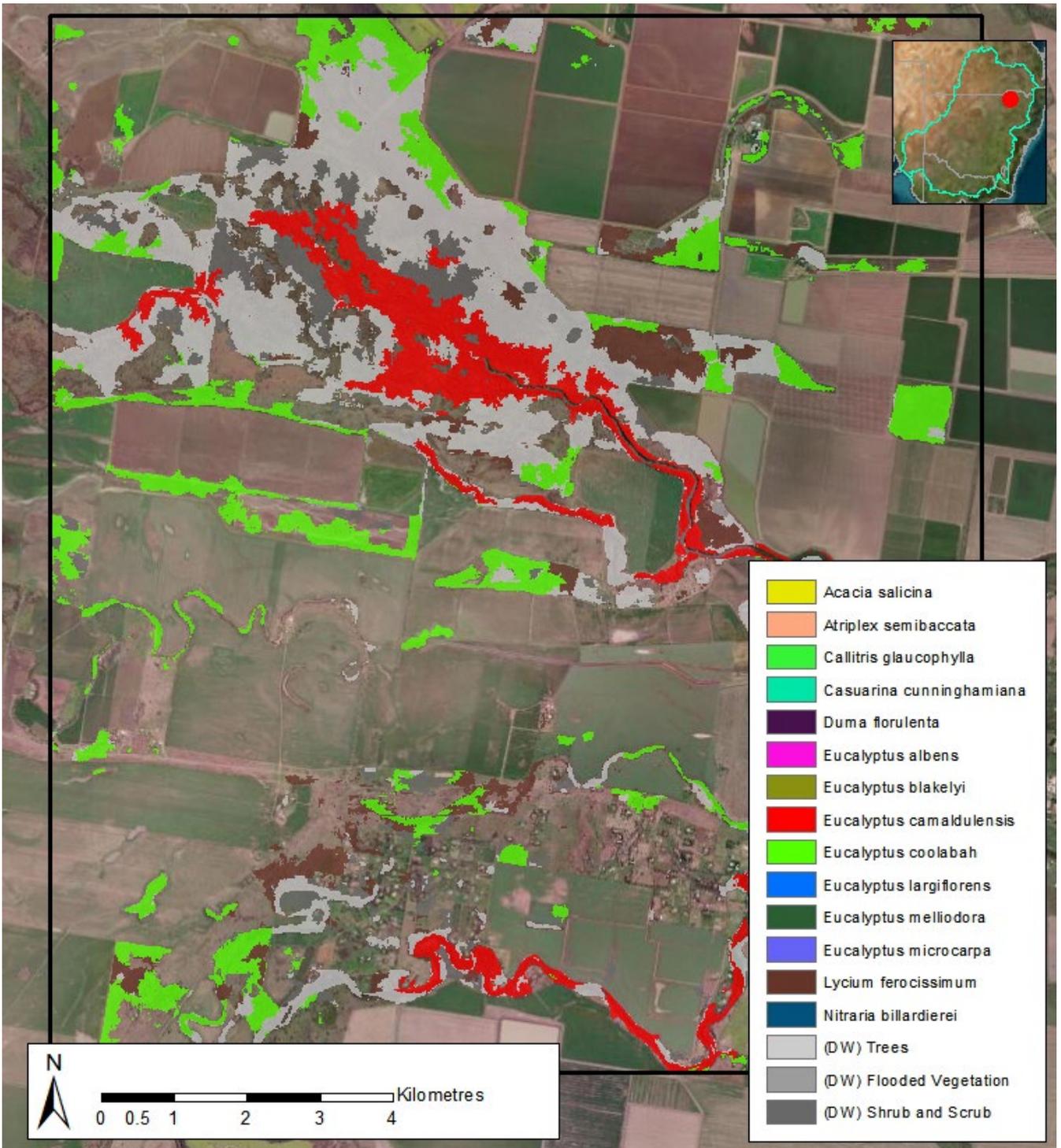


Figure 21. Desktop inspection site 5: Moree.

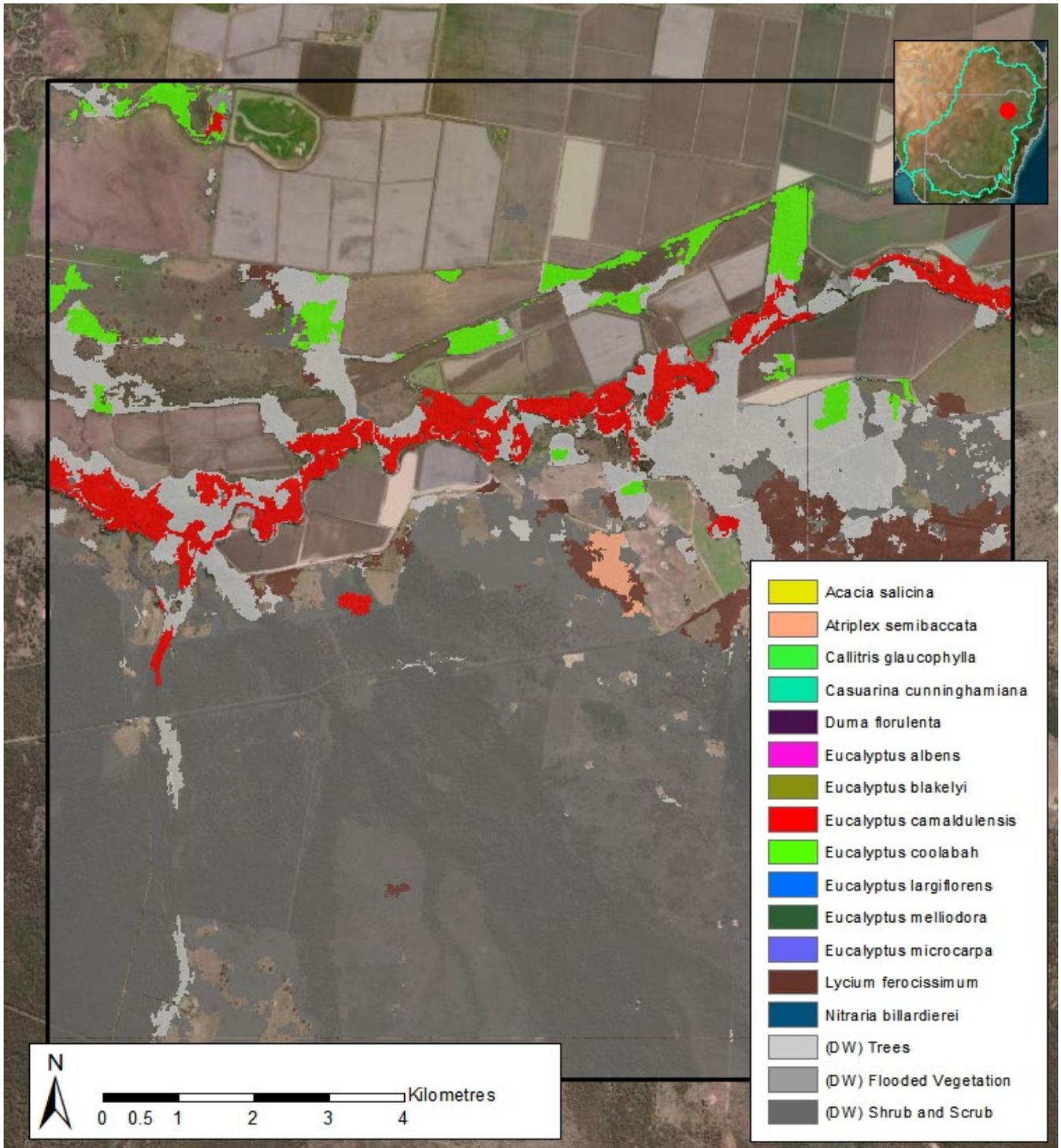


Figure 22. Desktop inspection site 6: Weewaa.

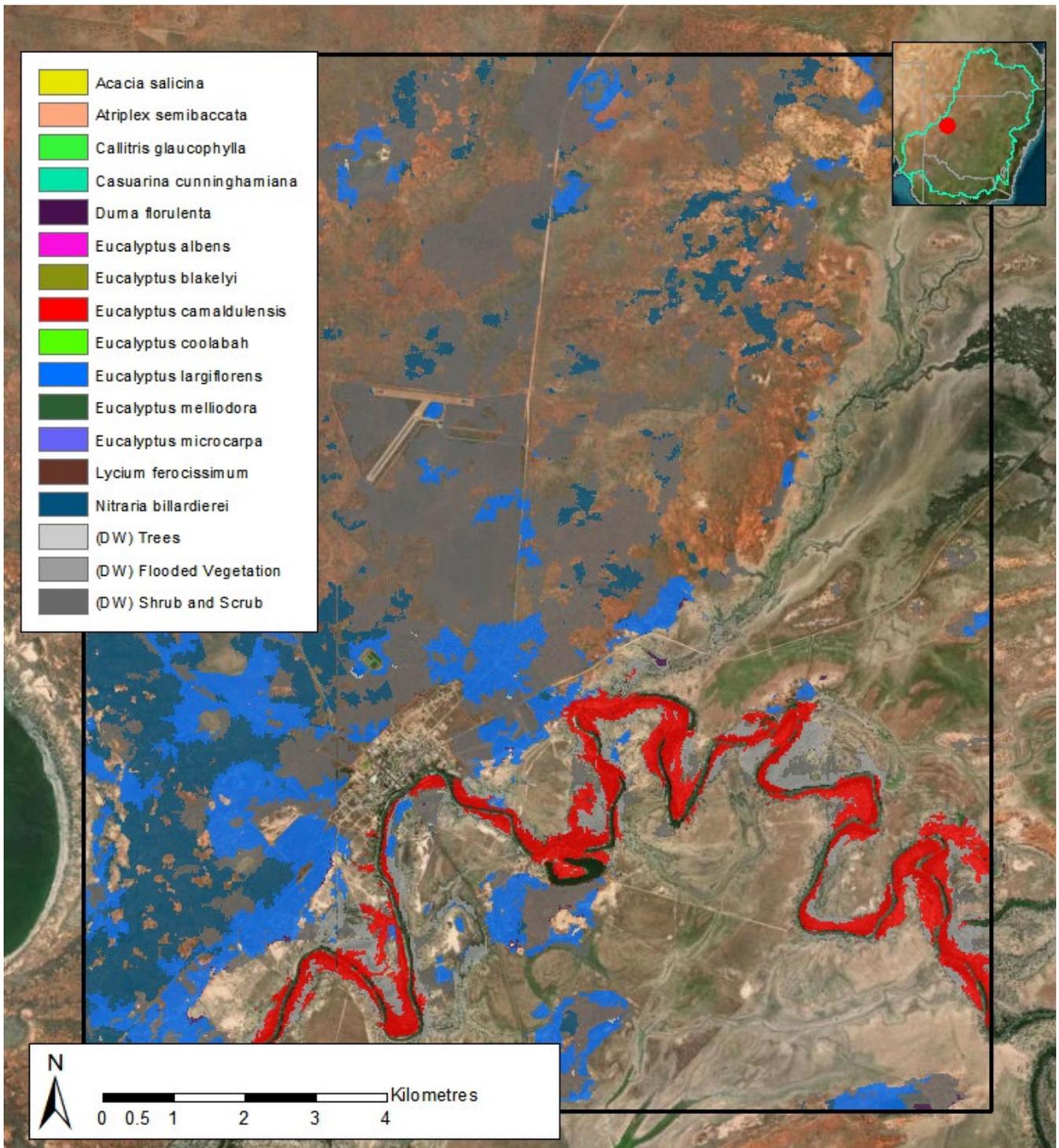


Figure 23. Desktop inspection site 7: Wilcannia.

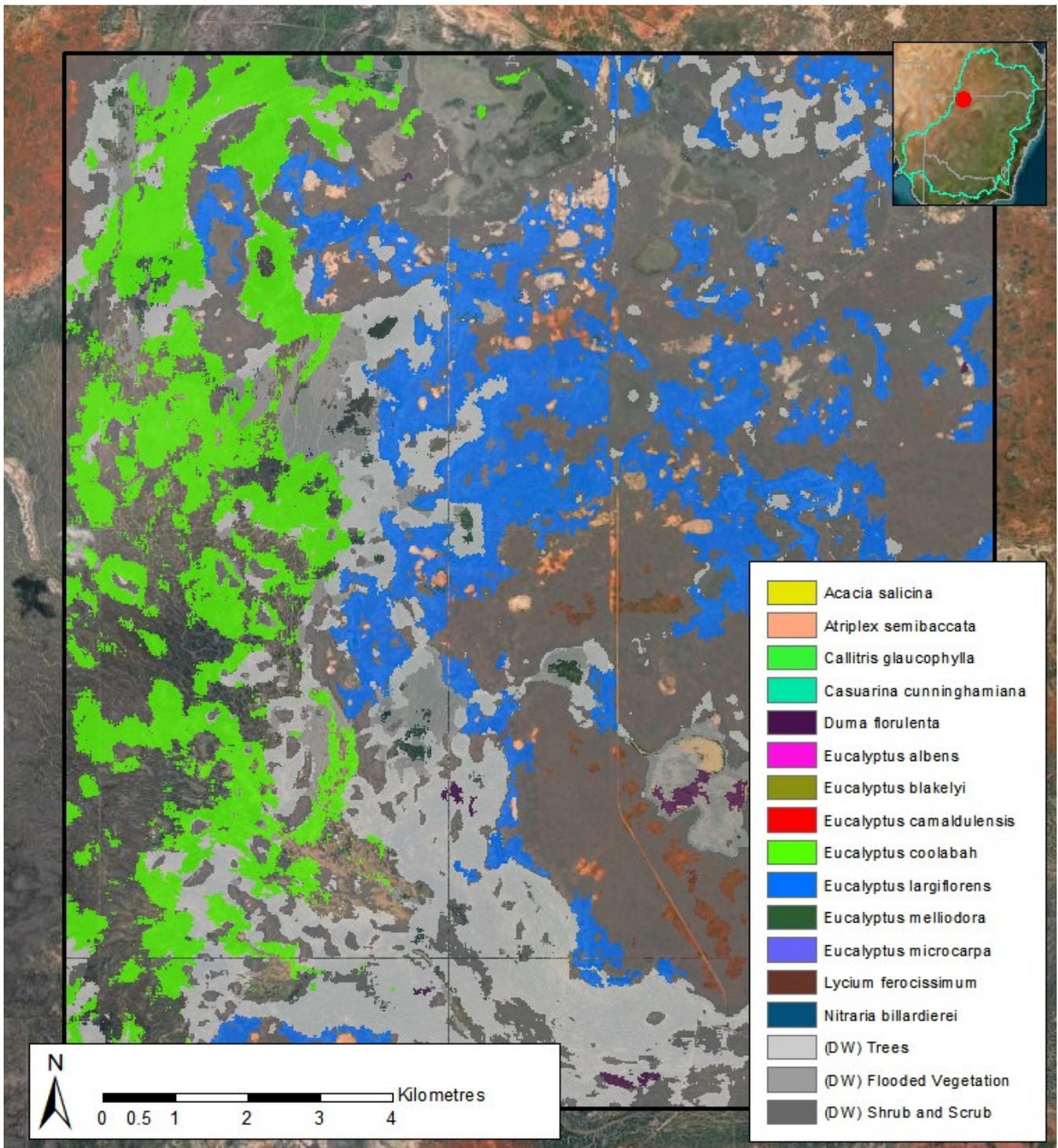


Figure 24. Desktop inspection site 8: Yantabulla.

5.4 Change detection demonstration

Here, a general example of change detection for 2021 and 2022 is presented, with the maps and detected change shown in Figure 25. Several small areas that were classified as *Eucalyptus microcarpa* in 2021 were then classified as *Eucalyptus camaldulensis* in 2022.

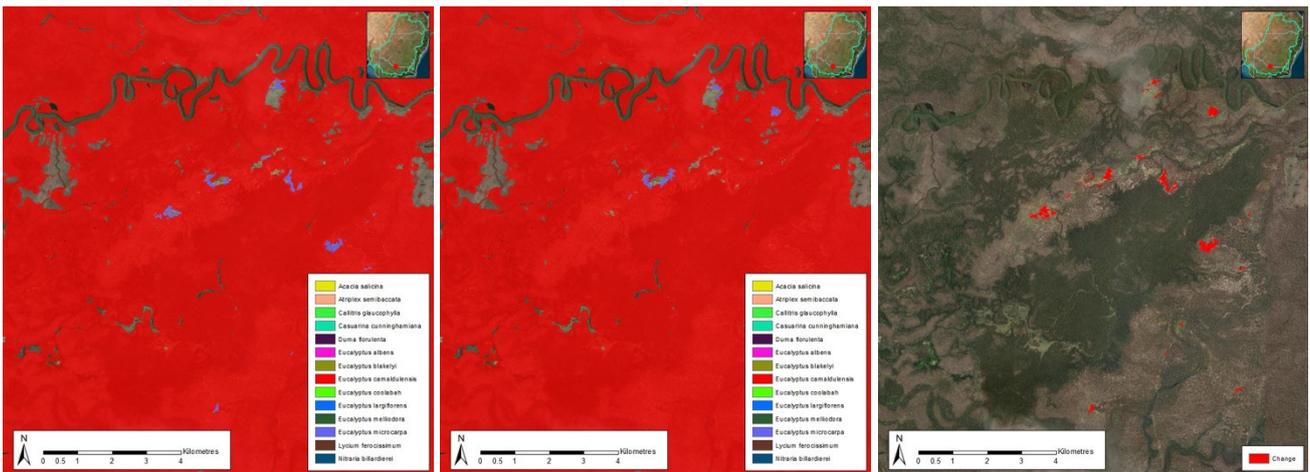


Figure 25. Change detection demonstration. Left: 2021 hard classification map. Middle: 2022 classification map. Right: Change in hard classification.

Zooming in on one of these changed areas, the confusion index for 2021 and 2022 is shown in Figure 26. The confusion indices show low confidence when separating these two classes in the 2021 data but higher confidence the following year (the 2022 data). More specifically, the model was more confident when the area was classified as *Eucalyptus camaldulensis* in 2022 but it was less confident when it classified the same area as *Eucalyptus microcarpa* in 2021. Considering the broader area (Figure 25), it is evident that this change occurs within a larger *Eucalyptus camaldulensis* patch. This suggests that the change is the result of a misclassification of the 2021 data. To be certain however, the user will need to apply additional knowledge (such as reviewing imagery of the area, local knowledge, or independent ground survey data) to conclude whether change has occurred. It is entirely possible to create simple apps that will allow experts to update the layers after desktop or field visual inspections, but this is outside the scope of this work.

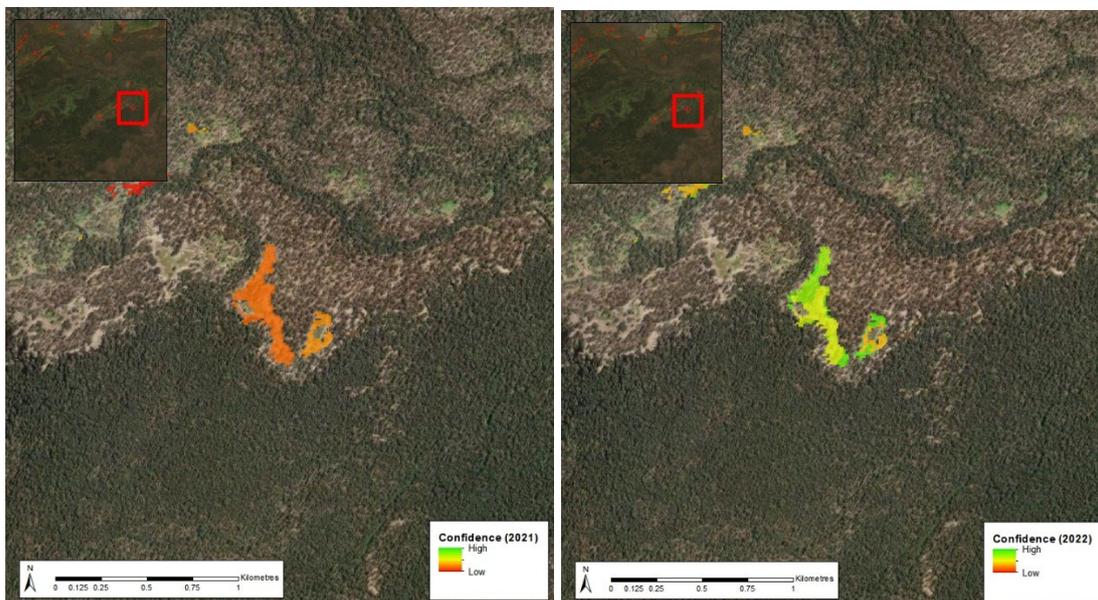


Figure 26. Confusion index for 2021 (left) and 2022 (right).

6 Discussion

This section, interprets and discusses the results from model training and validation, the species maps, and change detection demonstration.

6.1 Model training and validation

6.1.1 Preliminary and secondary modelling

The two model training phases produced multiple insights that informed the final model structure and have value for future modelling.

The project team found that having a high sample count for each class is critical. Sample counts of thousands per class produced a more accurate model than sample counts in the low hundreds. This was particularly evident when moving from sampling the temporal features in 2022 to sampling temporal features between 2019 and 2022, shown in the improvement in overall accuracy in Table 7. It was also evident when looking at annual and seasonal variability of species, for example, as shown in Figure 7; species with under 300 unique samples did not produce consistent annual feature distributions, making them difficult to accurately and reliably model.

Over the course of the model development, object-based sampling was chosen over pixel-based sampling because it produced a well-segmented classification map, without scattered pixel misclassifications. It has the downside of losing some independent samples due to grouping within objects compared to the pixel-based approach (demonstrated in Table 6), but was not found to be an issue once the number of samples was increased through temporal sampling.

While having a high number of samples is important, it was also found that removing low quality samples can improve the overall accuracy. Restricting the samples to those with the highest plot confidence score produced a higher overall accuracy score than keeping samples with plot confidence scores of 3-5 (demonstrated in Table 6). This indicated that the quality of the sample is important, and that it is possible to produce a better result with a smaller number of high-quality samples.

During the preliminary development phase, it was found that a conservative correlation threshold of 0.9 eliminated many of the temporal features of the model and produced an overly simplistic species mapping, which was more akin to a habitat map than a species map. While upping the correlation threshold to 0.99 resulted in keeping many more features (increasing the computational cost of the model), this produced a map that better reflected expected species distributions.

During the second round of development, removing species and features that varied strongly depending on the year improved the overall accuracy score for the model, suggesting that features and species with low annual variability in the feature space can be more accurately mapped than those that have high variability (demonstrated in Table 7). The removal of species that have high annual variability will also help with producing a stable map that can be used for annual change detection, as there will be fewer instances of model confusion.

6.1.2 Final model

The final model had an overall accuracy of 80.3%. The model produced consistently low errors of commission, which ranged between 6.4% to 24.1%, while errors of omission ranged between 4.8% and 48.6%. This is expected, as there are commonly trade-offs between commission and omission errors, and the model with lower commission errors was preferred by the MDBA team. This is because maps with low commission errors provide more confidence that the mapped species are truly present, with fewer false positives than maps with high commission errors. High performing species included *Nitraria billardierei* (commission = 11.8%, omission = 4.8%), *Callitris glaucophylla* (commission = 20.0%, omission = 7.7%), and *Casuarina cunninghamiana* (commission = 9.7%, omission = 19.1%). Environmental conditions, sampling size and the sampling subset are the likely drivers of success for these species. *Casuarina cunninghamiana* occurs in the tallest strata of the modelled species (see Appendix 2) with relatively high sample count and constrained distribution along the east of the Basin. *Callitris glaucophylla* occurs in the second tallest strata and is broadly sampled but has a concentrated collection within drier areas. *Nitraria billardierei* occurs in the third highest strata and only has a few samples with limited distribution. However, they are found to grow in strong saline saltmarsh conditions, which are unsuitable for many of the other mapped species.

Lower performing species were *Eucalyptus microcarpa*, *Duma florulenta*, and *Atriplex semibaccata*. Although all three had comparable rates of commission, they suffered high rates of omission at 31.6%, 45.6% and 48.6% respectively. Both *Atriplex semibaccata* and *Duma florulenta* are lower height strata species at 4 and 3 respectively. *Duma florulenta* was relatively lowly sampled compared to other species and suffered higher rates of confusion with *Eucalyptus largiflorens* and *Lycium ferocissimum*. *Atriplex semibaccata* also suffered high confusion with *Lycium ferocissimum*. *Eucalyptus microcarpa* was assigned to the tallest strata and was frequently misclassified as *Callitris glaucophylla*. This was likely due to both

species being sampled from dry vegetation areas, and the higher number of *Callitris glaucophylla* samples in the training dataset (2647 samples of *Callitris glaucophylla* to 716 samples of *Eucalyptus microcarpa*).

When reviewing the important features by rank (Table 11), it is notable that elevation is the most important feature. This is explained by the fact that elevation is a determinant of temperature, soil composition and water availability (particularly groundwater). It is a known driver of vegetation community and was found to be an important feature in the species models developed by both Cunningham et al. (2013) and Liang et al. (2020). The other features within the top 20 are also broad drivers of vegetation including soil characteristics, temperature, and topography. The favouring of environmental features in the model explains the higher performance of species that grow in specific environmental conditions, leading them to be more easily distinguished from other species.

In a broader assessment of the relative distribution of features by importance ranking, environmental features were the most important with biophysical characteristics from remote sensing data not emerging until the 40th Rank (demonstrated in Table 12 and Figure 11). Of the highest ranked biophysical features, the NIR, Blue, and SWIR bands appeared multiple times, along with the NDMI and NDWI indices (both relating to water).

6.2 Species mapping

For this project, the species mapping was provided as both a workflow and layered results. By delivering the workflow in a modular way, the MDBA will be able to further refine the species mapping process from sample aggregation through to data classification. This includes modifications to existing processing steps (e.g., the aggregation of plots), or the addition, removal, or merging of new species and/or sample data.

The layered results include the core likelihood layers (where values represent the likelihood of a pixel of belonging to that class) and a hard classification layer (a thematic label). It should be noted that this hard classification layer represents a generalised way to disseminate the classes. The Murray-Darling Basin is a complex environment with high temporal variability to both seasonal and annual weather patterns. Applying a one-size-fits-all thematic mapping to this environment will inevitably introduce compromise. The per-species likelihood layers offer an accessible, flexible dataset that can be tailored towards specific purposes.

The key species of interest in the Murray–Darling Basin are riparian/floodplain species, and these species were the primary focus of this work. Expansion beyond this riparian area, while attempted, resulted in some challenges. This can be seen in the distribution of species (Figure 12), which can be broken down into three main areas: the wetter eastern seaboard, inner wet drainage areas, and the inner dry areas. The inner wet drainage areas can be seen to shift among a range of species depending on region including *Eucalyptus camaldulensis* (south), *Eucalyptus coolabah* (north), *Eucalyptus largiflorens* (west), and *Nitraria billardiarei* (south west). *Eucalyptus melliodora* dominates along the eastern seaboard, which aligns to the distribution of its sampling. Moving inland, *Eucalyptus microcarpa* starts appearing. *Callitris glaucophylla* also has high area coverage across the map, dominating the classification prediction in drier areas. Although its distribution seems likely, its dominance does not. This is likely due to low presences of other sampled species in these areas (Figure 27), the broader sampling of *Callitris glaucophylla*, resulting in the model over-generalising dry vegetation as *Callitris glaucophylla*.

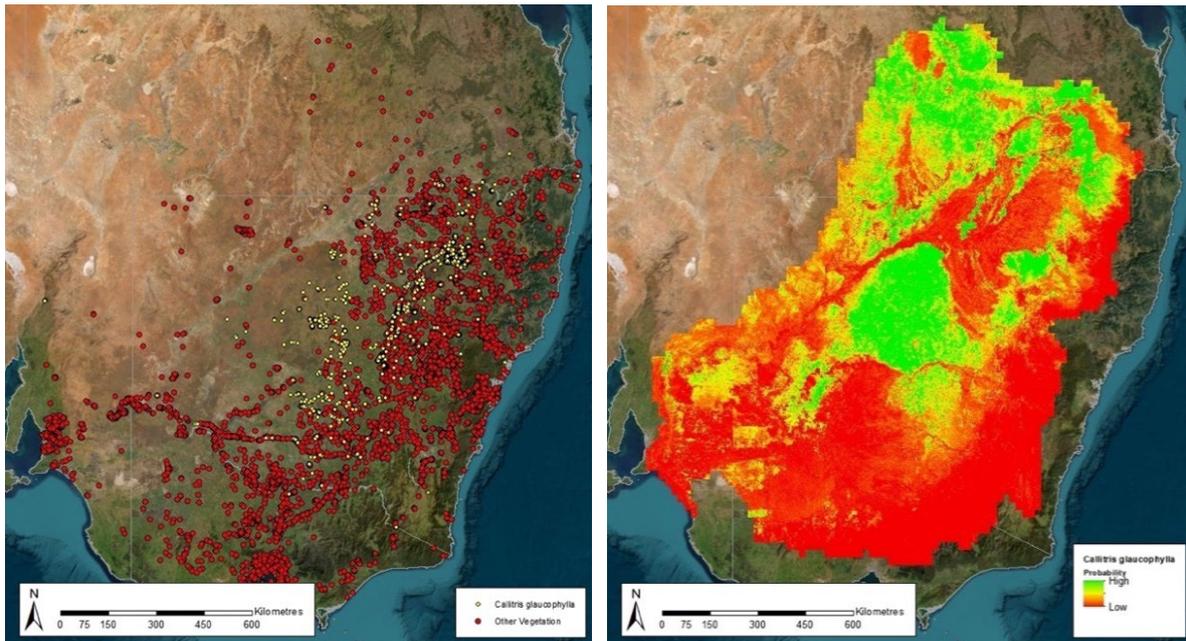


Figure 27. Left: *Callitris glaucophylla* samples (yellow) dominate large inland areas with few or no samples from all other species (red). Right: Likelihood for *Callitris glaucophylla*.

The over-generalisation of dry area vegetation to *Callitris glaucophylla* represents a significant source of classification error. This is further exacerbated by the model's very high confidence that dry vegetation represents *Callitris glaucophylla* (see Figure 27). From Table 8, the only confusion is with other inland woody species: *Eucalyptus albens* and *Eucalyptus microcarpa*. It is noted that the introduction of this error is a limitation of the validation approach: the confusion matrix is not well suited to identifying errors that result from a non-representative sample. As such, this error was not identified until the Basin-wide process was run for 2022. Future work could address this error by including a broader selection of dry woody tree species, with broader spatial distribution of sampling. To mitigate this error in the final hard classification, all areas classified as *Callitris glaucophylla* were downgraded to a broader "shrub and scrub" class.

Another challenge for this project was the unavailability of independent local-scale extent mapping that could be used for quantitative map-based validation. Using 20% of all data collected for this project as the validation set was sufficient for quantitatively assessing model performance but was not as valuable for assessing the quality of the classification map (for this the visual inspection of selected sites was relied upon). The most problematic aspect of the final approach is the limitation to point data, without any broader coverage or boundary information. Point data cannot be used to measure the level of noise in the map or identify segmentation artefacts. More generally, it is noted that validation data is limited to areas where samples were collected and may not represent broader distribution; this was the challenge for *Callitris glaucophylla*, as discussed above. Future establishment and expert surveying (desktop or ground) of ground sites will significantly strengthen the outcomes of future work on this dataset.

Desktop inspection was used to qualitatively assess the model. The visual inspection of selected sites revealed that the model generally reproduces expected vegetation distributions, such as the domination of Barmah Forest by *Eucalyptus camaldulensis* (Figure 17), and the presence of *Eucalyptus camaldulensis* along river banks (e.g. Figure 19, Figure 21, Figure 23).

The Basin-wide maps generated by this project were also compared to those from Cunningham et al. (2013). The similarity in the distribution of the three key species with those found by Cunningham et al. (2013) lends support to both models, noting that they were independently developed ten years apart, using different data and machine learning model design. The new approach improves on Cunningham et al. (2013) by accurately mapping many more species; for the species that both methods mapped, similar omission errors and much lower commission errors were produced (Table 10).

Finally, even though remote sensing features were not within the top 20 features of the final model, their inclusion was critical when it came to mapping. Environmental features were favoured by the model because they produced relatively low noise information that related to distribution of vegetation species. This contrasts with the contribution from remote sensing features, which captured higher local variation at the expense of greater noise. When working with low sample sizes in the preliminary modelling phase, it was found that removing the remote sensing features produced a higher accuracy (Table 6) but produced a map that over generalised large areas of the landscape into a single class. In response to this, the sampling

of temporal features was expanded from a single year (2022) to multiple years (2019, 2020, 2021, and 2022). The resulting model was able capture both the broader spatial environmental characteristics and local spatial variation in the final map, justifying the value of remote sensing data in addition to environmental data.

6.3 Change detection

During the project, the project team were unable to identify a specific known event (e.g. fire or deforestation in an area of known species extent) for the change detection demonstration. As such, a general example of the change detection capability was demonstrated (Section 5.4).

The key element of the project's approach to change detection is the use of the likelihood layers in addition to the hard classification. This adds more depth to how a change can be interpreted. The demonstration used a classification from 2021 and 2022. Areas that have changed class in the thematic map were highlighted. From there, the confusion index was calculated for both years. The demonstration showed that the difference in the confusion indices can help indicate which year the model has more certainty in the hard classification. This may help distinguish meaningful change: if the model was confident in both years, it may indicate that a real change has occurred.

More information, such as known areas of change, are needed to further calibrate and tailor the change detection approach. Future work is needed to explore if one or multiple approaches are needed for different kinds of change, for example, growth in extent of a mapped species, loss of a mapped species due to fire, or loss of a mapped species due to deforestation.

7 Conclusions

The project has successfully produced a system that the MDBA can use to run annual species mapping and report on changes to species extents as part of reporting against the key outcomes of the 2019 BWS. The benefit of the approach is that each component of the workflow can be further customised by the MDBA or another party, which will allow the method to be improved over time. The method also offers flexibility by generating the species likelihood layers, from which customised, fit-for-purpose, and accurate species extent maps can be derived.

The final machine learning model had an overall accuracy of 80.3% and produced similar species-level omission errors to the model developed by Cunningham et al. (2013) and lower commission errors. The final 2022 map showed similar distributions of the three key species (*Eucalyptus camaldulensis*, *Eucalyptus largiflorens* and *Eucalyptus coolabah*) to the maps produced by Cunningham et al. (2013), lending support to this new approach. The project has made a significant advancement on the method developed by Cunningham et al. (2013) in that it produced a method for annual mapping, which takes annual variability into account. As the method uses remote sensing data from multiple years during training, it is expected that the model will improve as more satellite imagery is collected by the two main remote sensing satellites used in this work: Sentinel-2 and Sentinel-1.

One challenge for this project was the availability of independent validation data. To this end, the project team made the following recommendations:

- Collate any existing independent local species maps to use as validation for species mapping in future development, and any future species mapping projects.
- Conduct further validation of the Basin-wide hard classification and likelihood layers, leveraging existing knowledge of stakeholders across the Basin.
- Refine the post-processing steps to produce a hard classification layer that meets the needs of the MDBA's stakeholders. This could involve updating the masking with additional datasets available to the MDBA or tweaking the likelihood thresholds on a per species basis.
- Identify known cases of change detection and validate the demonstrated approach developed in this project.

Another key lesson from this work is that ground survey data is necessary as the basis for any species extent model, but that it need not be the only approach. Specifically, data augmentation in the form of adding additional temporal sampling for known ground survey data massively boosted the number of samples available to the final machine learning model, which directly boosted the final overall accuracy. The collection of additional survey data was considered as part of this project, but was not pursued as this would have been a large expense for the collection of a few hundred points, when thousands are required to make a substantial improvement to model accuracy. As such, it is recommended that the MDBA work with other data collection initiatives to leverage any additional species presence or quadrat data that is collected within the Basin. By collaborating with teams that are collecting survey data for other reasons, the MDBA may be able to leverage additional data at much lower cost than through independent contracting of a ground survey.

Appendix

Appendix 1. Common attributes across all databases used in the project.

Attribute	Description
scientificname	The record's scientific name, which is composed of its genus name followed by its species name.
genus	The genus name of the record, extracted from the scientific name.
species	The species name of the record, extracted from the scientific name.
vernacularname	The common name used for this species. Data was cleaned to use a common set of vernacular names for this study. This may not be available for all records.
eventdate	The date the record was taken.
decimallongitude	The longitude of the record, measured in decimal degrees in WGS 84 (EPSG:4326).
decimallatitude	The latitude of the record, measured in decimal degrees in WGS 84 (EPSG: 4326).
coordinateuncertaintyinmeters	The uncertainty in the record's position recorded in meters.
geometry	Shapely Point geometry comprised of the longitude and latitude in GDA 94 (EPSG:4283).

Appendix 2. Hight strata used to determine species dominance and relative confidence for plots. Heights correspond to typical heights for grown members of the species. The tallest species are assigned to stratum 1; the shortest species are assigned to stratum 4.

Name	Growth Form	Height	Assigned Strata
<i>Casuarina cunninghamiana</i>	Woody (Tree)	Up to 35 m	1
<i>Eucalyptus albens</i>	Woody (Tree)	Between 15 to 25 m	1
<i>Eucalyptus blakelyi</i>	Woody (Tree)	Between 20 to 25 m	1
<i>Eucalyptus camaldulensis</i>	Woody (Tree)	Up to 20 m.	1
<i>Eucalyptus coolabah</i>	Woody (Tree)	Up to 20 m	1
<i>Eucalyptus largiflorens</i>	Woody (Tree)	Up to 20 m	1
<i>Eucalyptus melliodora</i>	Woody (Tree)	Up to 30 m	1
<i>Eucalyptus microcarpa</i>	Woody (Tree)	Up to 25 m	1
<i>Acacia salicina</i>	Woody (Tree/Shrub)	Up to 13.7 m	2
<i>Acacia stenophylla</i>	Woody (Tree/Shrub)	Between 4 to 20 m	2
<i>Callitris glaucophylla</i>	Woody (Tree)	Between 4 to 12 m	2
<i>Lycium ferocissimum</i>	Woody (Shrub)	Up to 5 m	2
<i>Atriplex nummularia</i>	Woody (Shrub)	Up to 3 m	3
<i>Chenopodium nitrariaceum</i>	Woody (Shrub)	Up to 2 m	3
<i>Duma florulenta</i>	Woody (Shrub)	Up to 2.5 m	3
<i>Nitraria billardierei</i>	Woody (Shrub)	Up to 2 m	3
<i>Atriplex semibaccata</i>	Herbaceous (Groundcover)	Between 0.4 to 0.8 m	4
<i>Atriplex suberecta</i>	Herbaceous (Groundcover)	Between 0.4 to 0.8 m	4

Bibliography

- Achanta, Radhakrishna, and Sabine Susstrunk. 2017. 'Superpixels and Polygons Using Simple Non-Iterative Clustering'. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4895–4904. IEEE. <https://doi.org/10.1109/CVPR.2017.520>.
- Aquatic Ecosystems Task Group. 2012. 'Aquatic Ecosystems Toolkit. Module 2. Interim Australian National Aquatic Ecosystem Classification Framework'. Canberra.
- Badola, Anushree, Hitendra Padalia, Mariana Belgiu, and Prabhakar Alok Verma. 2021. 'Tree Species Mapping in Tropical Forests Using Hyperspectral Remote Sensing and Machine Learning'. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 5421–24. IEEE. <https://doi.org/10.1109/IGARSS47720.2021.9553549>.
- Bolyn, Corentin, Philippe Lejeune, Adrien Michez, and Nicolas Latte. 2022. 'Mapping Tree Species Proportions from Satellite Imagery Using Spectral–Spatial Deep Learning'. *Remote Sensing of Environment* 280 (October): 113205. <https://doi.org/10.1016/j.rse.2022.113205>.
- Breiman, Leo. 2001. 'Random Forests'. *Machine Learning* 45 (1): 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Chowdhury, Md. Sharafat. 2024. 'Comparison of Accuracy and Reliability of Random Forest, Support Vector Machine, Artificial Neural Network and Maximum Likelihood Method in Land Use/Cover Classification of Urban Setting'. *Environmental Challenges* 14 (January): 100800. <https://doi.org/10.1016/j.envc.2023.100800>.
- Cunningham, SC, M White, P Griffioen, G Newell, and R Mac Nally. 2013. 'Mapping Floodplain Vegetation Types across the Murray-Darling Basin Using Remote Sensing'. Canberra.
- Fassnacht, Fabian Ewald, Hooman Latifi, Krzysztof Stereńczak, Aneta Modzelewska, Michael Lefsky, Lars T. Waser, Christoph Straub, and Aniruddha Ghosh. 2016. 'Review of Studies on Tree Species Classification from Remotely Sensed Data'. *Remote Sensing of Environment* 186 (December): 64–87. <https://doi.org/10.1016/j.rse.2016.08.013>.
- Gao, Bo-cai. 1996. 'NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space'. *Remote Sensing of Environment* 58 (3): 257–66. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3).
- Google. 2023. 'API Reference: Ee.Algorithms.Image.Segmentation.SNIC'. 6 October 2023. <https://developers.google.com/earth-engine/apidocs/ee-algorithms-image-segmentation-snic>.
- Grabska, Ewa, David Frantz, and Katarzyna Ostapowicz. 2020. 'Evaluation of Machine Learning Algorithms for Forest Stand Species Mapping Using Sentinel-2 Imagery and Environmental Data in the Polish Carpathians'. *Remote Sensing of Environment* 251 (December): 112103. <https://doi.org/10.1016/j.rse.2020.112103>.
- Grossmann, Emilie, Janet Ohmann, James Kagan, Heather May, and Matthew Gregory. 2010. 'Mapping Ecological Systems with a Random Forest Model: Tradeoffs between Errors and Bias'. *Gap Analysis Bulletin*, no. 17: 16–22.
- Liang, Wanwan, Mongi Abidi, Luis Carrasco, Jack McNelis, Liem Tran, Yingkui Li, and Jerome Grant. 2020. 'Mapping Vegetation at Species Level with High-Resolution Multispectral and Lidar Data Over a Large Spatial Area: A Case Study with Kudzu'. *Remote Sensing* 12 (4): 609. <https://doi.org/10.3390/rs12040609>.
- Lim, Joongbin, Kyoung-Min Kim, Eun-Hee Kim, and Ri Jin. 2020. 'Machine Learning for Tree Species Classification Using Sentinel-2 Spectral Information, Crown Texture, and Environmental Variables'. *Remote Sensing* 12 (12): 2049. <https://doi.org/10.3390/rs12122049>.
- Masemola, Cecilia, Moses Azong Cho, and Abel Ramoelo. 2020. 'Sentinel-2 Time Series Based Optimal Features and Time Window for Mapping Invasive Australian Native Acacia Species in KwaZulu Natal, South Africa'. *International Journal of Applied Earth Observation and Geoinformation* 93 (December): 102207. <https://doi.org/10.1016/j.jag.2020.102207>.
- McFeeters, S. K. 1996. 'The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features'. *International Journal of Remote Sensing* 17 (7): 1425–32. <https://doi.org/10.1080/01431169608948714>.
- Mohammadpour, Pegah, Domingos Xavier Viegas, and Carlos Viegas. 2022. 'Vegetation Mapping with Random Forest Using Sentinel 2 and GLCM Texture Feature—A Case Study for Lousã Region, Portugal'. *Remote Sensing* 14 (18): 4585. <https://doi.org/10.3390/rs14184585>.
- Mokany, Karel, David Peel, Rocio Ponce Reyes, Scott Foster, Andrew Freebairn, Danial Stratford, Heather McGinness, Brenton Zampatti, Sam Nicol, and Paul McInerney. 2023. 'Modelling and Mapping Habitat for Key Species in the Murray-Darling Basin'. Canberra.
- Murray–Darling Basin Authority. 2019. 'Basin-Wide Environmental Watering Strategy'. Canberra.
- Pu, Ruiliang. 2021. 'Mapping Tree Species Using Advanced Remote Sensing Technologies: A State-of-the-Art Review and Perspective'. *Journal of Remote Sensing* 2021 (January). <https://doi.org/10.34133/2021/9812624>.
- Qi, J., A. Chehbouni, A.R. Huete, Y.H. Kerr, and S. Sorooshian. 1994. 'A Modified Soil Adjusted Vegetation Index'. *Remote Sensing of Environment* 48 (2): 119–26. [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1).
- Science Education through Earth Observation for High Schools. n.d. 'Classification Algorithms and Methods'. Accessed 2 January 2024. <https://seos-project.eu/classification/classification-c01-p05.html>.
- Shang, Xiao, and Laurie A. Chisholm. 2014. 'Classification of Australian Native Forest Species Using Hyperspectral Remote Sensing and Machine-Learning Classification Algorithms'. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (6): 2481–89. <https://doi.org/10.1109/JSTARS.2013.2282166>.
- Yang, Lingbo, Limin Wang, Ghali Abdullahi Abubakar, and Jingfeng Huang. 2021. 'High-Resolution Rice Mapping Based on SNIC Segmentation and Multi-Source Remote Sensing Images'. *Remote Sensing* 13 (6): 1148. <https://doi.org/10.3390/rs13061148>.