Northern Australia Environmental Resources Hub

National Environmental Science Programme

Remote sensing of gamba grass in northern Australia: An assessment of methods and guidelines for use

Report

Shaun R Levick, Yuri Shendryk, Stephanie Johnson, Samantha Setterfield and Natalie Rossiter-Rachor

CSIRO

CHARLES DARWIN UNIVERSITY AUSTRALIA

THE UNIVERSITY OF WESTERN AUSTRALIA

This report should be cited as:

Levick SR, Shendryk Y, Johnson S, Setterfield S and Rossiter-Rachor N. (2021). *Remote sensing of gamba grass in northern Australia: An assessment of methods and guidelines for use.* CSIRO, Canberra.

Cover photograph: Sentinel-1 (ESA/ATG medialab).

This report is available for download from the Northern Australia Environmental Resources (NAER) Hub website at nespnorthern.edu.au

The Hub is supported through funding from the Australian Government's National Environmental Science Program (NESP). The NESP NAER Hub is hosted by Charles Darwin University.

ISBN 978-1-922684-33-2

October, 2021

Printed by UniPrint

Contents

List of figures

List of tables

Acronyms

BA..... balanced accuracy

CHIME..... Copernicus Hyperspectral Imaging Mission

DESIS..... DLR Earth Sensing Imaging Spectrometer

DLR ..... German Aerospace Centre

EnMAP..... Environmental Monitoring and Analysis Program

ERF..... Emissions Reduction Fund

ESA ..... European Space Agency

ISRO..... Indian Space Research Organization

MODIS..... Moderate Resolution Imaging Spectroradiometer

NASA ..... National Aeronautical Science Administration

NDSI..... Normalised Difference Spectral Index

NIR..... near-infrared

NISAR ..... NASA-ISRO Synthetic Aperture Radar

PAN..... panchromatic

RADAR..... RAdio Detection And Ranging

RGB..... red, green, blue

SAR..... synthetic aperture radar

SBG..... Surface Biology and Geology

SFPS ..... step-forward predictor selection

SRTM ..... Shuttle Radar Topography Mission

SWIR ..... short-wave infrared

VHR..... very high resolution

VIIRS……Visible Infrared Imaging Radiometer Suite

VNIR..... visible/near-infrared

Executive summary

The work described in this report is part of Project 2.3 'Weed invasion, fire and ecosystem failure: catchment scale scenario modelling to improve planning and management' of the Northern Australia Environmental Resources Hub of the Australian Government's National Environmental Science Program. A core objective of this project was to work in consultation with the Department of Agriculture, Water and the Environment to develop, if possible, a remote sensing approach at a scale, reliability and cost suitable for mapping and monitoring gamba grass to meet the requirements of the Carbon Credits (*Carbon Farming Initiative—Savanna Fire Management—Emissions Avoidance*) Methodology Determination 2018.

This report explores a range of current remote sensing approaches to mapping and monitoring gamba grass presence and density for their suitability for monitoring in the Emissions Reduction Fund (ERF) methodologies. It begins with an evaluation of current remote sensing missions and identifies important developments with relevance to gamba grass and other high-biomass grass mapping for the ERF. After considering the scale, reliability and cost of the remote sensing methods, a multi-scaled remote sensing approach that integrates optical and RADAR sensing was suggested as being the most likely to achieve the mapping and monitoring required for the ERF. Although a large range of sensors and satellites exist, each with their own strengths and weaknesses, our recommendation is that research should focus on the fusion of Sentinel-1 (synthetic aperture radar) and Sentinel-2 (multi-spectral) data streams from the European Space Agency (ESA). Used in conjunction, these constellations provided imagery at high spatial (10–20 m) and temporal (6–10 days) resolutions, facilitating mapping of both structural and spectral vegetation properties through different seasons. Both data streams are freely available under the Copernicus Programme and are readily accessible through multiple distributed storage and cloud-based processing platforms.

The suggested approach was tested by mapping the distribution of gamba grass in the Batchelor/Litchfield/Adelaide River region of the Northern Territory. A two-tiered strategy was adopted using a combination of i) commercial, high-resolution, 16-band, multi-spectral WorldView-3 imagery for model development at local scales (100–200 km^2), and ii) the fusion of Sentinel-1/-2 time-series imagery for mapping at larger region scales (5,000 km^2) with machine learning. The local-scale WorldView-3 model, trained with on-ground field data, produced excellent results and returned a balanced accuracy of 91% for gamba grass occurrence at the 0.3 m spatial resolution. These high-resolution outputs were aggregated to 20 m spatial resolution and used as calibration/validation data to teach another machine-learning model (gradient-boosted regression) based on Sentinel-1 and Sentinel-2 data. The regional model also performed well, returning a balanced accuracy of 82%. Validation of model outputs across the broader region, which encompasses multiple land uses, showed that gamba grass presence was underestimated when it occurred beneath the canopy of riparian trees, and was overestimated in some instances along the boundaries of wetlands, where confusion with native wetland grass occurred. The presence of isolated gamba grass plants (1 m^2) was detectable from WorldView-3, but patches needed to be in the order of 10–20 m^2 before detection from Sentinel-1/-2 was possible. Future modelling efforts could refine these results by making greater use of temporal signatures, coupled with deep learning approaches that can better account for time-series trends (e.g. recurrent neural networks). In coming years, the launch of multiple spaceborne hyperspectral missions (NASA's SBG, ESA's CHIME and the German Aerospace Centre's EnMAP) will facilitate greater delineation among similarly structured species.

1. Requirements for the mapping and monitoring of gamba grass in the Emissions Reduction Fund

Regions across northern Australia are under increasing threat from a range of high-biomass invasive grasses (Setterfield et al. 2018). These species result in ecosystem degradation, habitat loss and species decline in the ecosystems they invade (Rossiter-Rachor et al. 2009, Brooks et al. 2010), primarily due to their impact on fire, and have been declared as a Key Threatening Process under the *Environment Protection and Biodiversity Conservation Act 1999* (Setterfield et al. 2018). Of these serious invasive weeds, gamba grass (*Andropogon gayanus*) is of particular concern. Gamba grass outcompetes native grass species, replacing the understorey with dense, tall monocultures up to 4 m high (Ens et al. 2015). The increased above-ground biomass (fuel loads) in invaded ecosystems result in fire intensities up to eight times greater than in native savannas (Rossiter et al. 2003, Setterfield et al. 2010), leading to substantial tree decline in invaded savannas that are repeatedly burnt (Ferdinands et al. 2006). Gamba grass invasion and gamba grass-fuelled fires pose a significant risk to carbon stocks and fluxes (Setterfield et al. 2008), and carbon abatement projects across northern Australia (Adams and Setterfield 2013).

The revised *Carbon Farming Initiative—Savanna Fire Management—Emissions Avoidance Methodology Determination 2018* includes formal requirements for savanna emissions avoidance project areas to exclude areas that contain a relevant weed species (14.2(b) (i)), defined in the accompanying technical guidance document as a weed species which materially affects fire dynamics. The only species currently listed as a relevant weed species is *Andropogon gayanus* (gamba grass).

The Department of Agriculture, Water and the Environment's *Savanna Technical Guidance Document* describes the mapping and monitoring instructions for relevant weed species (Section 8.1). Of relevance for this project is that:

- A project must submit a vegetation map, mapped at 250×250 m pixel scale.
- The project area must not include an area of land that contains a relevant weed species. If a relevant weed species occurs in a project area, then the project is an ineligible offset project.
- Projects must be monitored for the presence of relevant weed species as required by the savanna fire management determinations.

A key consideration for developing a cost-effective remote sensing method for gamba-grass mapping is the size of the properties that require monitoring and the diversity of vegetation types that will be assessed for gamba grass presence. Current Emissions Reduction Fund (ERF) properties are spread across the Northern Territory, Queensland and Western Australia, and range in size from 300 to 30,000 km^2, with the largest being Kakadu National Park and the West Arnhem Land Fire Abatement (WALFA) project (Figure 1). Accuracy of monitoring methods therefore needs to be assessed across a broad range of vegetation types. The complexity of distinguishing gamba grass (Figure 2) from other high-biomass grasses that may be structurally and spectrally similar, such as *Megathyrsus maximus* (guinea grass) and *Cenchrus* species, also needs to be considered.

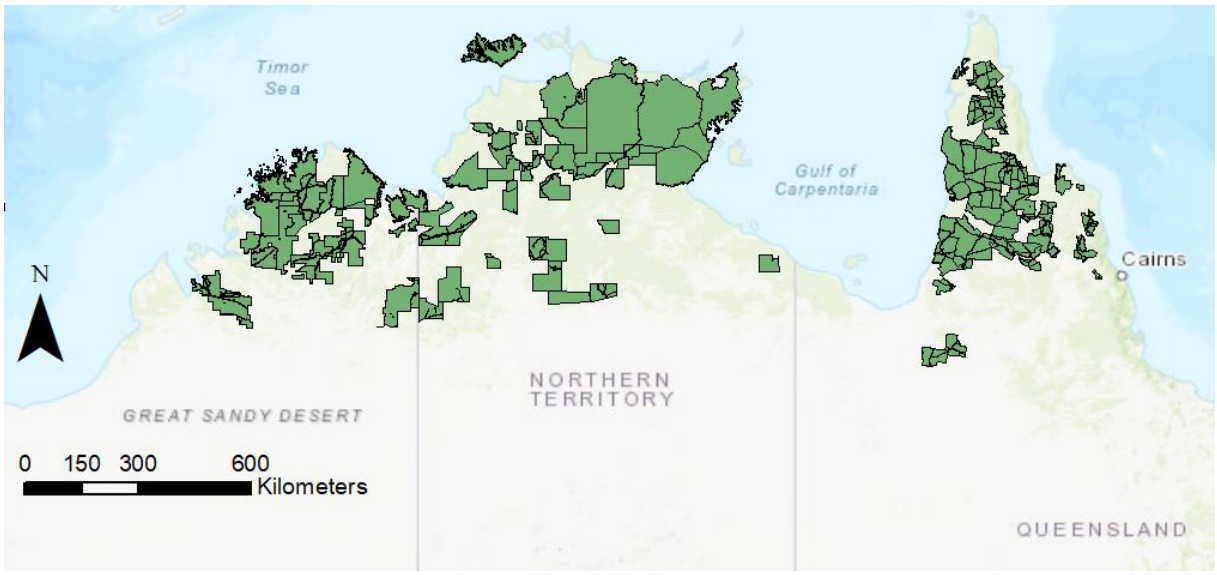


Figure 1. Map of properties enrolled in the Emissions Reduction Fund in 2018. Individual projects range in size from 300 to 30,000 km^2.

Figure 2. Dense stand of gamba grass in the Batchelor region of the Northern Territory, Australia. Photo: Natalie Rossiter-Rachor.

2. Satellite remote sensing requirements for high-biomass grass mapping

Savanna grasses are notoriously difficult to map from space, due to their small individual plant size, co-occurrence with trees and seasonally dynamic phenology (Skidmore et al. 2010, Ali et al. 2016). These issues are exacerbated in Australia's savannas by very high fire frequency, with fires occurring, on average, in two out of every three years (Beringer et al. 2015). Objects without distinct structural or spectral properties are difficult to classify in remotely sensed imagery (Blaschke et al. 2014, Bradley 2014). However high-biomass invasive grasses such as gamba grass are so structurally large (Figure 2) in comparison to most smaller-statured native grasses (Rossiter-Rachor et al. 2008, Setterfield et al. 2010) that detection from space within ERF project areas is potentially feasible. Furthermore, the phenology of gamba grass differs from many of its native neighbours, with plants staying green (photosynthetically active) longer into the dry season and curing much later than native grasses (Rossiter-Rachor et al. 2009, Setterfield et al. 2013). As such, there is potential to leverage spectral properties to distinguish gamba grass from native grasses, provided that imagery of a suitable resolution can be acquired at regular intervals at key times of the year, such as the start and end of the dry season.

We are fortunate to have a large range of satellite programs and sensors available for Earth observation research, and the number of sensors has increased exponentially in recent years (Belward and Skoien 2015, Gorelick et al. 2017, Lewis et al. 2017). Each program has different objectives, so sensor and satellite configurations differ in their core characteristics across a spectrum of spatial, temporal, spectral and radiometric domains. No one satellite can excel in all four of these dimensions, so trade-offs need to be made in selecting the sensor, or combination of sensors, most suitable to the mapping problem at hand.

2.1 Spatial, temporal, spectral and radiometric considerations

Spatial resolution refers to the pixel size of an image covering the Earth's surface and is key to determining the size of the minimum mapping unit. Sensors offering higher spatial resolution have the advantage of being able to distinguish smaller features in landscapes, but often sacrifice temporal and radiometric resolution. Commercial satellites, such as WorldView-3 and Pleiades, offer spatial resolutions of <0.5 m, but imagery needs to be tasked and, in most cases, only 1–2 images will be collected for a particular location per year.

Temporal resolution is also referred to as the revisit time and denotes the frequency of image collections – how often each location on Earth is imaged. Higher temporal frequencies are advantageous in cloudy environments (by increasing the probability of attaining a cloud-free image) and are useful in phenological studies aimed at identifying trends in photosynthetic activity. MODIS is an example of a satellite with a high temporal resolution, collecting imagery twice per day for a given location. The trade-off is its coarse spatial resolution (500 m).

Spectral resolution specifies the number of spectral bands and the range of wavelengths in which a sensor can collect reflected radiance.

Radiometric resolution refers to the sensitivity of a sensor to the magnitude of the electromagnetic energy received.

Trade-offs between these factors cannot be avoided when sensors are engineered (Turner et al. 2003). Sensors need to have a small instantaneous field-of-view (IFOV) to record images with a high spatial resolution. However, as the pixel sizes become smaller and the area within the IFOV becomes smaller, the amount of energy that can be detected also decreases. Reduced energy means reduced radiometric resolution, and it becomes harder to detect small differences in reflectance. To increase the radiometric resolution without reducing the spatial resolution, it would be necessary to broaden the wavelength range for a particular channel/band, but this would have consequences for the spectral resolution. Coarser spatial resolution sensors allow for improved radiometric and spectral resolutions, and are often also associated with better temporal resolutions, as satellites with broader swathes can orbit the globe faster (Haldar 2013).

2.2 Current satellite options for remote sensing of gamba grass in the ERF context

A large range of current satellite options could potentially be used for mapping gamba grass in ERF properties (Table 1). However, preliminary investigations of available imagery and existing literature indicate that no single sensor can provide the necessary information because of the constraints described above. Despite its physically large structure relative to native grasses, the presence of gamba grass amidst native trees and other grasses will be challenging to detect. As such, it is necessary to leverage different types of sensors for the detection of subtle spectral and structural signals.

Very high-resolution (VHR) sensors (<1 m) like WorldView-3 (Figure 3), Pleiades, and KOMPSAT-3 have the best chance of detection, but they are costly and better suited to specific site studies than broad-scale property mapping. As such, while VHR imagery can contribute information at specific sites, a more feasible option for the ERF property scale is the Sentinel program from the European Space Agency (ESA), which encompasses high- and medium-resolution optical imagers (Sentinel-2 and Sentinel-3) and a C-band RADAR system (Sentinel-1; Table 1). Incorporating synthetic aperture radar (Sentinel-1) is beneficial as it can penetrate the cloud layer that is prominent over northern Australia for most of the wet season, and its spatial resolution is closely matched to the 10 m optical sampling of Sentinel-2 (Figure 4). Furthermore, by analysing the full time-series from these satellites (images every 5–10 days) it is possible to map features that are not evident in single time steps (Main et al. 2016).

*Table 1. Current satellite options with potential to provide information on vegetation structure and composition. * denotes missions that are not yet operational, ** denotes missions experiencing extended delays, – denotes no cost to user.*

	Agency	Platform	Type	Bands or polarisations	Spatial resolution (m)	Revisit interval (days)	Cost
	NASA	MODIS	Multi-spectral	36 (620–14,385 nm)	250, 500, 1,000	1	–
	NASA	Landsat 8	Multi-spectral	11 (435–12,510 nm)	15, 30, 100	16	–
	ESA	Sentinel-1	RADAR	C-band (dual HH+HV or VV+VH)	5–100	5-10	–
	ESA	Sentinel-2	Multi-spectral	13 (443–2,190 nm)	10, 20, 60	5-10	–
	ESA	Sentinel-3	Multi-spectral	11 (555–10,850 nm)	500–1,000	1–2	–
	ESA	Proba-V	Multi-spectral	4 (415–176 nm)	100	2	–
	Planet	RapidEye	Multi-spectral	5 (440–850 nm)	5	1–5	$
	Planet	PlanetScope	Multi-spectral	4 (440–850 nm)	3.125	1	$
	DigitalGlobe	WorldView-3	Multi-spectral	17 (450–2,365 nm)	0.31, 1.24, 3.7	task	$$$
	JAXA	ALOS-2	RADAR	L-band (quad HH+HV+VH+VV)	10–100	14	$$
	DLR	TanDEM-X	RADAR	X-band (quad HH+HV+VH+VV)	1–18	11	$$
	AirBus	Pleiades	Multi-spectral	5 (470–940 nm)	0.5, 2	task	$$$
	KARI	KOMPSAT-3	Multi-spectral	5 (450–900 nm)	0.7, 2.8	task	$$
	SurreySpace	NovaSAR	RADAR	S-band (tri HH, VV, HV or VH)	6–30	1–4	$
*	NASA	GEDI	LiDAR	HOMER – 1064 nm	25	–	–
*	CSA	RADARSAT constellation	RADAR	C-band (quad HH+HV+VH+VV)	1–100	4	$
*	ESA	BIOMASS	RADAR	P-band (quad HH+HV+VH+VV)	50–200	30	–
**	DLR	EnMAP	Hyperspectral	230 (420–2,450 nm)	30	1–27	–

ALOS = Advanced Land Observing Satellite; BIOMASS = Biomass Mission for Carbon Assessment; CSA = Canadian Space Agency; DLR = German Aerospace Centre; EnMAP = Environmental Monitoring and Analysis Program; ESA = European Space Agency; GEDI = Global Ecosystem Dynamics Investigation; HOMER = High Output Maximum Efficiency Resonator; JAXA = Japan Aerospace Exploration Agency; KARI = Korea Aerospace Research Institute; KOMPSAT = Korea Multi-Purpose Satellite; LiDAR = Light Detection And Ranging; MODIS = Moderate Resolution Imaging Spectroradiometer; NASA = National Aeronautical Science Administration; RADAR = Radio Detection And Ranging; SAR = synthetic aperture radar.

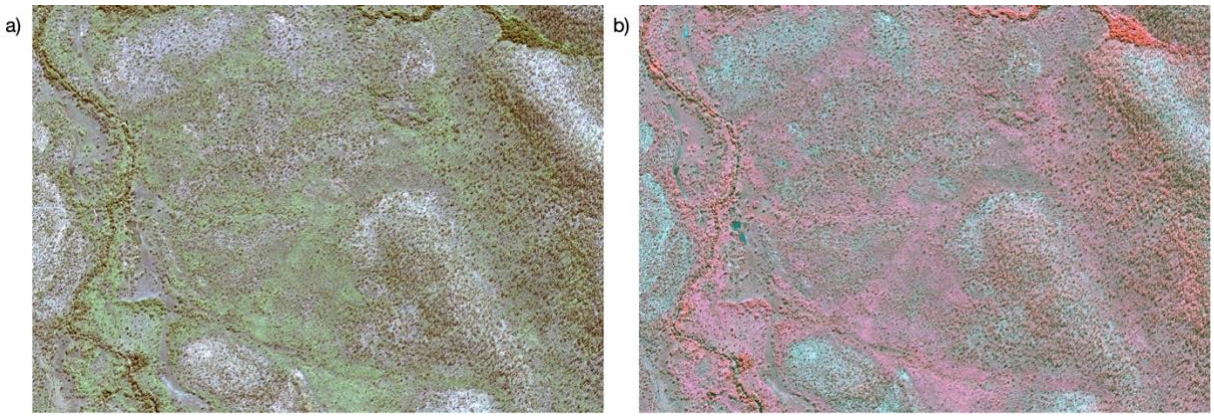


Figure 3. Examples of high spatial resolution imagery obtained from WorldView-3, showing (a) a true-colour composite and (b) a false-colour composite. Very high-resolution imagery such as this has the advantage of being able to delineate individual trees and create reliable endmembers for calibration data and spectral unmixing of coarser-resolution imagery. Gamba grass patches as small as 1 m^2 are identifiable.

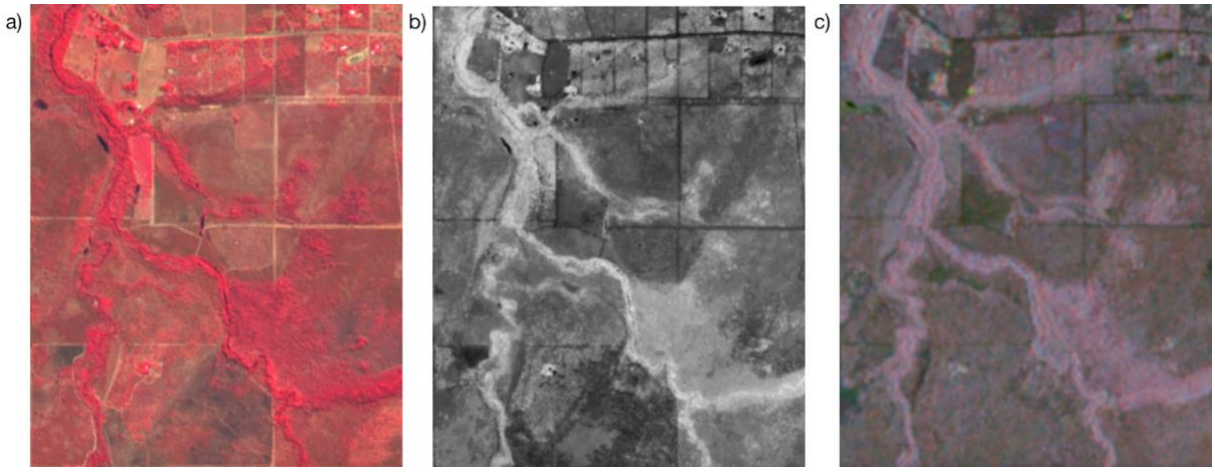


Figure 4. Recent imagery collected over the Batchelor area, Northern Territory, Australia, from the European Space Agency Sentinel constellations – (a) false-colour composite from Sentinel-2, (b) Normalised Difference Vegetation Index (NDVI) derived from Sentinel-2, and (c) a RADAR backscatter composite from Sentinel-1. Gamba grass patches need to be 10–20 m^2 for detection.

The Sentinel-1 and Sentinel-2 constellations are powerful in their own right, but by fusing the two together it is possible to draw on their complementary strengths – cloud penetration, vertical structure sensitivity, spectral properties and reduced revisit time (Clerici et al. 2017, Schmidt et al. 2018). This concept could be taken a step further by introducing another independent sensor of similar spatial and spectral properties, such as NASA's Landsat 8, thereby creating a virtual constellation with even higher temporal resolution (Li and Roy 2017), for key months at least, to target periods when native grasses are curing but gamba grass remains photosynthetically active. Recent developments in the ESA's Copernicus program have overcome some of temporal resolution trade-offs by moving in the direction of satellite constellations (Torres et al. 2012). A good example of this is the Sentinel-2 program, which currently comprises two satellites – Sentinel-2a and Sentinel-2b (Drusch et al. 2012). The two satellites travel the same track but are positioned opposite to each other in orbit. Sentinel-2a revisits a given location every 10 days – as does Sentinel-2b – but since they sit opposite each other in orbit, that given location is imaged every 5–6 days. As such, the constellation approach allows for an increase in temporal resolution without sacrificing spatial, spectral or radiometric resolution. In time, Sentinel-2c and Sentinel-2d will strengthen the program, adding redundancy and ensuring longevity.

2.2.1 Legacy satellite missions

The lifespan of satellite programs is difficult to predict, but recent missions have been very successful and there is a large degree of redundancy built into current global Earth-monitoring programs. A number of older sensors could still add value to the gamba-grass mapping challenge, but their longevity is questionable. NASA's MODIS, for example, has been a cornerstone of Earth observation since its launch in 2000 (Pettorelli et al. 2005). It has greatly exceeded its design life of six years and continues to provide imagery twice daily across the globe. Although MODIS is still widely used by the fire community in Australia for its burnt-area products (Hill et al. 2006, Maier 2010), we would not recommend developing new mapping initiatives based upon its lifespan. If very high temporal (daily) and greater spectral resolution is required over large areas, then emphasis should shift to Sentinel-3 and VIIRS which will largely fill this niche in the years ahead (Donlon et al. 2012, Wooster et al. 2012, Xiong et al. 2014).

For exploration of historical land cover and land cover change, the Landsat program is unparalleled (Wulder et al. 2018). Landsat imagery dating back to 1972 provides valuable insight into how ecosystem structure has changed over time, and is widely used across the globe for forest and water resource monitoring (Pflugmacher et al. 2012, Wulder et al. 2012, Hansen et al. 2013). In the context of high-biomass grass mapping in northern Australia, imagery from Landsat 5, 7 and 8 provide monthly time-series of land surface reflectance at 30 m spatial resolution. Although this archive provides useful background information on historic trends in land cover, from roughly the period when gamba grass was trialled as pasture grass in the Northern Territory, we did not consider analysis of longer-term historical dynamics to be essential for the current mapping challenge.

2.3 A multi-scaled approach for large-area mapping of gamba grass

Mapping high-biomass grasses in ERF properties requires the classification of gamba grass presence at the satellite pixel level based on spectral and structural features. These pixel presence/absence estimates can then be aggregated to the 250 m scale required for reporting of cover/density. The vast majority of land cover classification approaches from remotely sensed imagery are supervised procedures and require calibration (training) data composed of reference data of known status (Wulder et al. 2018). Machine learning – the use of statistical computer algorithms that improve automatically through experience and exposure to data (Jordan and Mitchell 2015) – holds much potential for improving classification accuracies of remotely sensed imagery. However, the limiting factor in leveraging artificial intelligence is typically the availability of sufficient training data for model calibration and validation.

To overcome this constraint, we adopted a multi-scaled approach whereby we first developed a high-resolution gamba grass cover map at a relatively small spatial scale (200 km^2), using machine learning to train a model based on high spatial and spectral resolution commercial imagery (WorldView-3) from field data. We then used the resulting cover map as training data for input into a second-order machine-learning model, which used open-access Sentinel-1 and Sentinel-2 imagery as input variables, to generate predications at much large spatial scales (5,000 km^2).

3. Leveraging commercial satellite imagery and gradient boosting for fine-scale gamba-grass mapping

The core results from the section below have been published as follows:

Shendryk, Y., Rossiter-Rachor, N. A., Setterfield, S. A., and Levick, S. R. (2020). Leveraging high-resolution satellite imagery and gradient boosting for invasive weed mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 4443–4450. https://doi.org/10.1109/JSTARS.2020.3013663

While both airborne hyperspectral and VHR satellite multi-spectral sensors can generate imagery with similarly high spatial resolutions (≤0.5 m), the latter are limited by spectral resolution, with current sensors usually providing 4-band imagery in red, green, blue (RGB) and near-infrared (NIR) wavelengths. At present, the most advanced, publicly available VHR satellite sensor is WorldView-3, which has 17 bands (including the panchromatic [PAN] band) in visible/near-infrared (VNIR) and short-wave infrared (SWIR) wavelengths (450–2,365 nm) (Table 1). The benefit of these additional spectral bands of VHR satellite multi-spectral sensors for mapping invasive weeds has not been well established, but it has been previously reported that imagery with additional spectral bands beyond RGB+NIR wavelengths did not improve the discrimination of invasive weeds (Marshall et al. 2012, Robinson et al. 2016). While textural features extracted from VHR satellite imagery have been previously found to capture the components of vegetation structure (Wood et al. 2012), there has been limited research into the benefits of textural features for classifying invasive weeds. The aim of this component of our study was to test the suitability of VHR WorldView-3 imagery for mapping the presence of gamba grass at high resolution at local property scales (100 km^2).

3.1 High-resolution mapping approach

To evaluate the sensitivity of WorldView-3 imagery in discriminating gamba grass, we tested the utility of the following input data: i) spectral bands, ii) textural features, iii) normalised difference spectral indices (NDSIs) and iv) all predictors in combination. We also investigated whether the additional 12 spectral bands that WorldView-3 offers beyond standard RGB+NIR wavelengths improves the discrimination of gamba grass. We employed an extreme gradient boosting (XGBoost) classification algorithm (Chen and Guestrin 2016) to take advantage of the high data dimensionality. To date, the most popular classification algorithms for mapping invasive weeds are random forest (Lawrence et al. 2006), maximum likelihood (Yang and Everitt 2010) and spectral angle mapper (Lass et al. 2005). However, decision tree boosting algorithms have shown good potential in multiple classification benchmarks (Briem et al. 2002, Olsony et al. 2018). The main advantages of XGBoost include: i) high computational speed due to parallel processing of data; ii) generally better performance in comparison to other decision tree-based models (if hyperparameters are tuned properly); iii) non-reliance on missing value imputation, scaling and normalisation of the input data; and iv) in-built regularisation terms that can be used to control the complexity of the model and avoid overfitting.

3.1.1 Study area

The 200 km^2 area of interest was located near the township of Batchelor, approximately 100 km south of Darwin, in the Northern Territory. The study area is largely under private ownership for pastoral lease or semirural development, with other significant areas owned by local communities or under government ownership. The area experiences a tropical climate with distinct wet and dry seasons. The annual rainfall is 1,535 mm, with the heaviest falls occurring during November to April (BOM 2020). The major vegetation type is savanna woodland dominated by *Eucalyptus miniata* and *E. tetrodonta*, with a herbaceous understorey dominated by native annual grass species, such as *Sorghum* spp., and perennial species such as *Heteropogon contortus* and *Alloteropsis semialata*, in addition to invasive perennial species including *Andropogon gayanus* (gamba grass). While gamba grass can be structurally similar to some *Sorghum* spp., gamba grass has a different phenological cycle to the native savanna grasses (Setterfield et al. 2013) and remains tall (up to 4 m) and photosynthetically active into the dry season (April–May); by this time, the native grasses have senesced (usually March–April) and 'collapsed' to form a low (~0.5–1 m) grass layer. Consequently, the early dry season is the best seasonal time to accurately detect gamba grass, when it is distinct and clearly visible in the landscape.

3.1.2 Field measurements for model training and validation

Field calibration/validation measurements were collected in multiple surveys conducted in March/April and July/August/September 2019. A Leica GS16 GNSS smart antenna and CS20 controller were used in conjunction with Leica's precise point positioning (PPP) service to enable centimetre accuracy of field capture. In the first survey, circular plots (radius = 3 m) were mapped with a homogeneous representation of: (1) live (green) gamba grass, (2) burnt gamba grass, (3) herbicide-sprayed gamba grass, (4) senesced *Sorghum* spp., (5) burnt *Sorghum* spp., and (6) other live and senescing native grass species. In the second survey, circular plots were only classed as gamba or non-gamba due to phenological changes that occur as the dry season progresses. Additional classes were defined after the fieldwork by delineating homogeneous circular plots (r = 3 m) of (7) trees and (8) waterbodies in WorldView-3 imagery. For the purpose of this study, samples were grouped into gamba-grass classes (i.e. 1, 2 and 3) and non-gamba-grass classes (i.e. 4, 5, 6, 7 and 8), resulting in an imbalanced (~1:2) training dataset of 187 and 355 samples respectively.

3.1.3 WorldView-3 imagery acquisition and processing

The WorldView-3 satellite was tasked to acquire imagery in April 2019 and two scenes were collected within 42 seconds of each other on 11 April under cloud-free conditions. April is the transition between wet and dry seasons, and was selected to task the satellite imagery acquisition because at that time *Sorghum* spp. and other native grasses have commenced senescence and browning in colour, while gamba grass remains photosynthetically active and green, and as such, is recognisable from the air (Petty et al. 2012). Each scene consisted of PAN, VNIR and SWIR bands in a wavelength range of 400–2,365 nm. The satellite imagery was provided as an Ortho Ready 2A product, which was radiometrically corrected to ground reflectance, and projected to a plane using UTM 52S projection. Radiometric correction was accomplished using the atmospheric compensation (AComp) algorithm developed by DigitalGlobe for their WorldView sensor series imagery.

As a first pre-processing step, both VNIR and SWIR bands were pan-sharpened to 0.3 m resolution based on the PAN band and using Zhang's (2002) algorithm implemented in PCI Geomatica software. Then, pan-sharpened VNIR and SWIR bands, as well as the PAN band, were orthorectified using rational polynomial coefficients and a 30 m resolution SRTM-derived digital elevation model using PCI Geomatica OrthoEngine. Finally, PAN, VNIR and SWIR spectral bands were stacked and merged into a single mosaic with 0.3 m spatial resolution (Figure 5). From the PAN band, 12 textural features were derived using a filter window of 3×3 pixels (Baraldi and Parmiggiani 1995). Previous research suggests that using a filter window of 3×3 pixels results in the highest accuracy of vegetation parameter estimation in VHR satellite imagery (Zhou et al. 2017). Similarly, from VNIR and SWIR bands, 120 NDSIs were derived in succession from Coastal Blue to SWIR-8 spectral bands as follows:

$$NDSI_{(i,j)} = (R_i - R_j)/(R_i + R_j)$$

where R is the spectral reflectance, and *i* and *j* are numbers indicating the wavelengths (nm). Each NDSI is denoted as a combination of three-letter acronyms, for example, NDSI(COA,SW8) is denoted as COASW8. Finally, we calculated predictor variables from WorldView3-derived spectral bands, textural features and NDSIs. For this, we extracted statistics in 17×17 pixel windows (~26 m^2) to match the areas of field measurements (3 m radius) for each field sampling plot, resulting in a total of 1,036 predictor variables.

3.1.4 Extreme gradient boosting

We used an XGBoost algorithm to try and differentiate gamba grass from other vegetated and non-vegetated areas. XGBoost is an ensemble learning method that combines the predictive power of multiple linear models or decision trees using a boosting algorithm (Chen and Guestrin 2016). In boosting, a decision tree or linear regression that improves the model most is added to an ensemble at each iteration until the set number of estimators (i.e. *n*_estimators) has been achieved. In contrast to bagging techniques such as random forest, in which trees are grown to their maximum extent, boosting makes use of shallow trees with fewer splits. The training data were shuffled and split into 'train' (75% of the data) and 'test' (25% of the data) sets. Then, the XGBoost classifier with a binary logistic loss function and a tree booster was used to predict the presence of gamba grass in two stages. First, a randomised search on hyperparameters was performed using the 'train' dataset with a stratified five-fold cross-validation (cv) (training:testing ratio of 75:25). Six hyperparameters were optimised: *max_depth* in a range from 1 to 11, with an increment of 1; *learning_rate* of 0.001, 0.01, 0.1, 0.5, 1 and 2; *subsample* in a range from 0.2 to 1.0, with an increment of 0.1; *min_child_weight* in a range from 1 to 21, with an increment of 1; *gamma* of 0, 0.25, 0.5 and 1; and *n_estimators* of 100, 500 and 1,000. We used expert knowledge to specify reasonable ranges and increments of the hyperparameters. The best model according to a cv score was used to extract the importance of predictor variables in terms of gain, which is the relative contribution of the corresponding predictor to the model (Bergstra et al. 2012). Second, using predictors ranked according to their importance, another XGBoost classification using step-forward predictor selection (SFPS) was performed (Miller 2002). In each iteration, the predictor that previously best improved the model performance was added until addition of new predictors did not improve the performance in terms of a cv score. Eight XGBoost models were trained and optimised using predictor variables extracted from: (1) four spectral bands (bands 2, 3, 5 and 7); (2) eight spectral bands (bands 1 to 8); (3) 16 spectral bands

(bands 1 to 16); (4) textural features derived from the PAN band; (5) six NDSIs derived from four spectral bands (bands 2, 3, 5 and 7); (6) 28 NDSIs derived from eight spectral bands (bands 1 to 8); (7) 120 NDSIs derived from 16 spectral bands (bands 1 to 16); and (8) all predictor variables in combination.

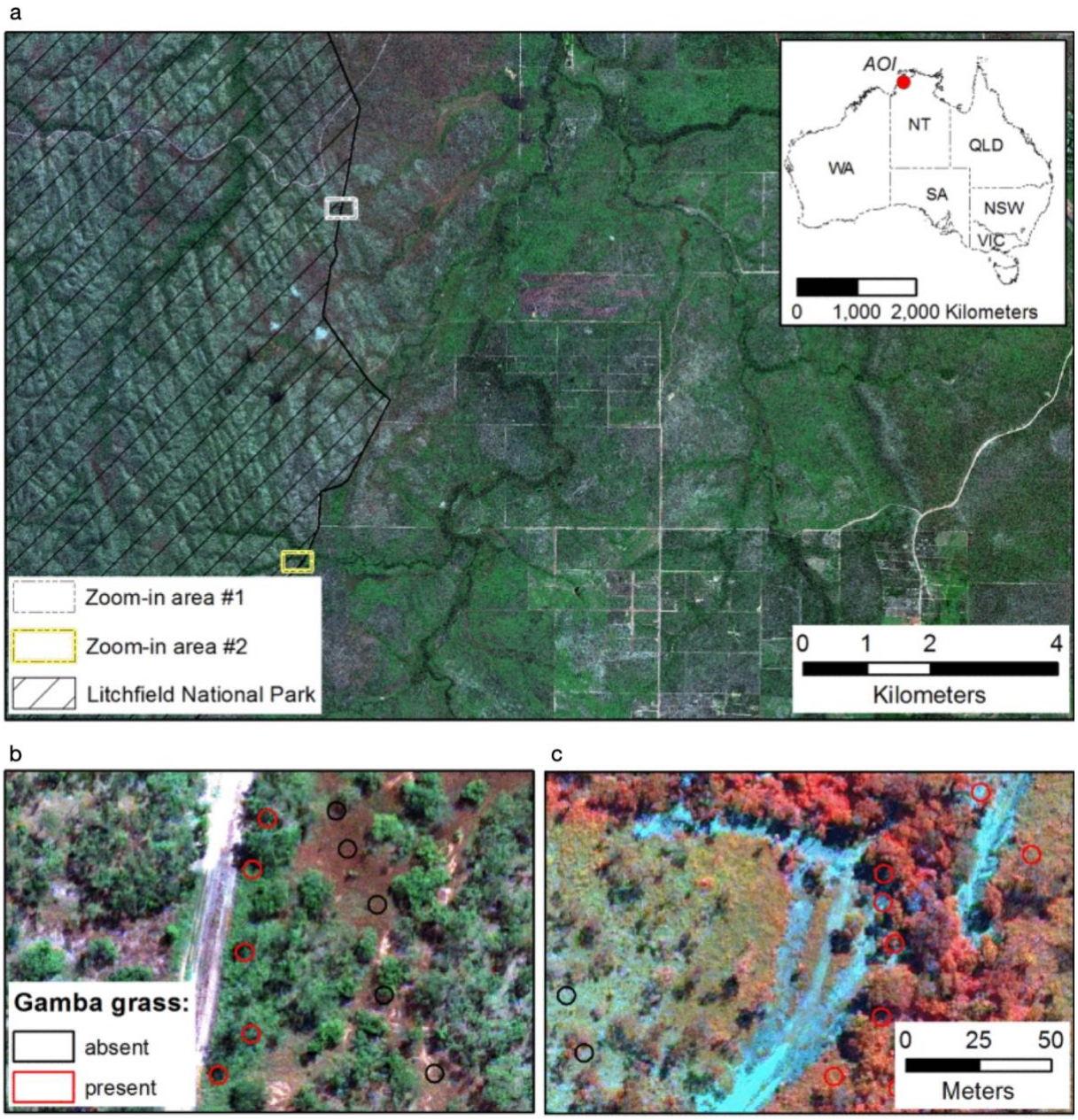


Figure 5. High-resolution imagery tasked for the projects from the WorldView-3 satellite. (a) True-colour composite showing the boundary between the Batchelor rural area and Litchfield National Park. Zoomed-in views show (b) true-colour (5,3,1) and (c) false-colour (8,4,1) composite examples with field validation points overlaid.

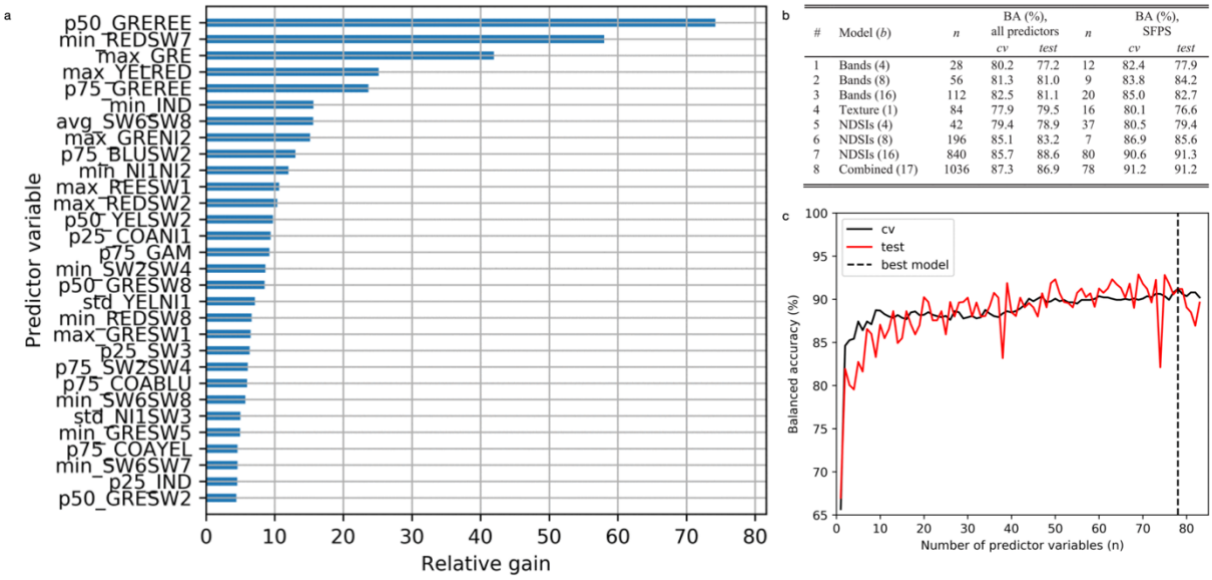
Multiple recent studies of VHR satellite imagery used object-based image analysis (OBIA) prior to classification for the detection of invasive weeds (Bradley 2014, Alvarez-Taboada et al. 2017). While OBIA has previously been shown to improve accuracies of classification compared with per-pixel classification (Levick and Rogers 2011, Whiteside et al. 2011), it usually relies on segmentation algorithms that are difficult to validate and correct. In this study, to use the power of OBIA approaches without relying on segmentation algorithms, we trained XGBoost models using predictor variables calculated within field-measured areas (~28.3 m^2) and applied it to predictor variables calculated within a 17×17 pixel neighbourhood area (~26 m^2). To optimise and evaluate our models during cross-validation and test stages, we used the balanced accuracy (BA) metric, which avoids inflated performance estimates on imbalanced datasets. It is the macro average of recall scores per class or, equivalently, raw accuracy where each sample is weighted according to the inverse prevalence of its true class, and is defined as:

$$BA = \left(\frac{\frac{TP}{P} + \frac{TN}{N}}{2} \right)$$

where TP is a true positive (correctly classified as positive), TN is a true negative (correctly classified as negative), P is a positive and N is a negative.

3.2 High-resolution mapping of gamba grass presence

The 30 most important predictors for the model trained using all predictor variables in combination is shown in Figure 6. Here, importance is expressed in terms of a relative gain, which is the relative contribution of the corresponding predictor to the model. The spectral bands identified by predictor importance as offering the greatest capacity for discrimination of gamba grass were those covering 510–745 nm and 2,185–2,365 nm wavelength ranges. The SFPS procedure generally resulted in an improved performance of the XGBoost classifier, with a BA increase on a 'test' set of up to 4.3% (Model 8 in Figure 6b), which is in line with previous findings (Robinson et al. 2016). The SFPS also reduced the number of necessary predictors for best performance by between 57% (Model 1) and 96% (Model 6). Although the combined use of all predictors (Model 8) in the SFPS procedure led to a BA increase from 86.9% to 91.2%, it was still no better than that of the NDSIs-derived model (Model 7). In contrast to previous studies, additional spectral bands did improve classification accuracy by 4.8% when going from a 4-band (Model 1) to a 16-band setup (Model 3). The improved performance became even more pronounced in an NDSIs scenario with a BA improvement between Model 5 and Model 7 of 11.9%. While the model trained using textural features only (Model 4) showed the worst performance with a BA of 76.6%, it was still comparable to that of Model 1 utilising RGB+NIR bands (BA of 77.9%). For Model 8, only three predictors (p50_GREREE, min_REDSW7 and max_GRE) were necessary to generate gamba grass presence classification with BA = 85%. Cross-validated BA stopped improving after using the top 78 predictors (Figure 6c). Nonetheless, the BA improvement between models trained using between 10 and 83 predictor variables could be considered marginal. The final map of gamba grass presence at 0.3 m spatial resolution generated using a model with hyperparameters from Model 8, trained using all training data (i.e. 'train' and 'test' sets combined) and the top 78 predictor variables, is shown in Figure 7.



#	Model (*b*)	*n*	BA (%), all predictors *cv*	BA (%), all predictors *test*	*n*	BA (%), SFPS *cv*	BA (%), SFPS *test*
1	Bands (4)	28	80.2	77.2	12	82.4	77.9
2	Bands (8)	56	81.3	81.0	9	83.8	84.2
3	Bands (16)	112	82.5	81.1	20	85.0	82.7
4	Texture (1)	84	77.9	79.5	16	80.1	76.6
5	NDSIs (4)	42	79.4	78.9	37	80.5	79.4
6	NDSIs (8)	196	85.1	83.2	7	86.9	85.6
7	NDSIs (16)	840	85.7	88.6	80	90.6	91.3
8	Combined (17)	1036	87.3	86.9	78	91.2	91.2

Figure 6. (a) Ranked importance of 30 predictor variables for the model trained using all predictor variables in combination (from highest to lowest): (b) Balanced accuracy of gamba grass presence classification using all predictors and predictors selected using step-forward predictor selection (SFPS); (c) SFPS for Model 8 using all predictor variables in combination.

Our results demonstrate that VHR WorldView-3 imagery can be used to differentiate gamba grass from other vegetated and non-vegetated areas, with accuracies of up to 91.3%. This methodology is technically scalable to larger areas as it relies exclusively on readily accessible VHR satellite imagery. This study is a significant advancement for stakeholders, as the accuracy of gamba-grass mapping is sufficient to inform landscape management in northern Australia. WorldView-3-derived NDSIs provided more separability than individual spectral bands or textural features when classifying gamba grass. Additional WorldView-3 VNIR bands (i.e. Coastal, Yellow, Red Edge and NIR-2) provided a 6.2% increase in classification accuracy, while the addition of SWIR bands improved classification accuracy by another 5.7%. However, given the spectral and structural differences of different vegetation types, this result might not be applicable when mapping other weed species (Marshall et al. 2012, Robinson et al. 2016). Interestingly, textural features extracted from the PAN band provided a satisfactory classification result with BA of 76.6%, which is relevant when considering the use of single-band sensors (e.g. WorldView-1) for mapping invasive weeds. An occasional misclassification of gamba grass presence occurred in areas shaded by tree crowns. These were characterised by change in the spectral shape, and were occasionally similar in terms of spectral and NDSI signal to unshaded gamba grass. This problem could be alleviated by collecting additional training samples in shaded areas to aid further discrimination.

While we expect VHR satellite imagery to be more accurate than medium-resolution satellite imagery in mapping invasive weeds (Matongera et al. 2017), numerous studies have also achieved accuracies of up to 90% when using medium-resolution satellite imagery for invasive weed mapping (Evangelista et al. 2009, Gavier-Pizarro et al. 2012). While being able to cover large areas and usually available for free, medium-resolution satellite imagery requires extensive field surveys at multiple time steps to achieve the above-mentioned accuracies. As such, we decided to use the mapping outputs from the VHR WorldView-3 analysis to inform medium-resolution satellite imagery to upscale the extent of gamba-grass mapping in time and space.

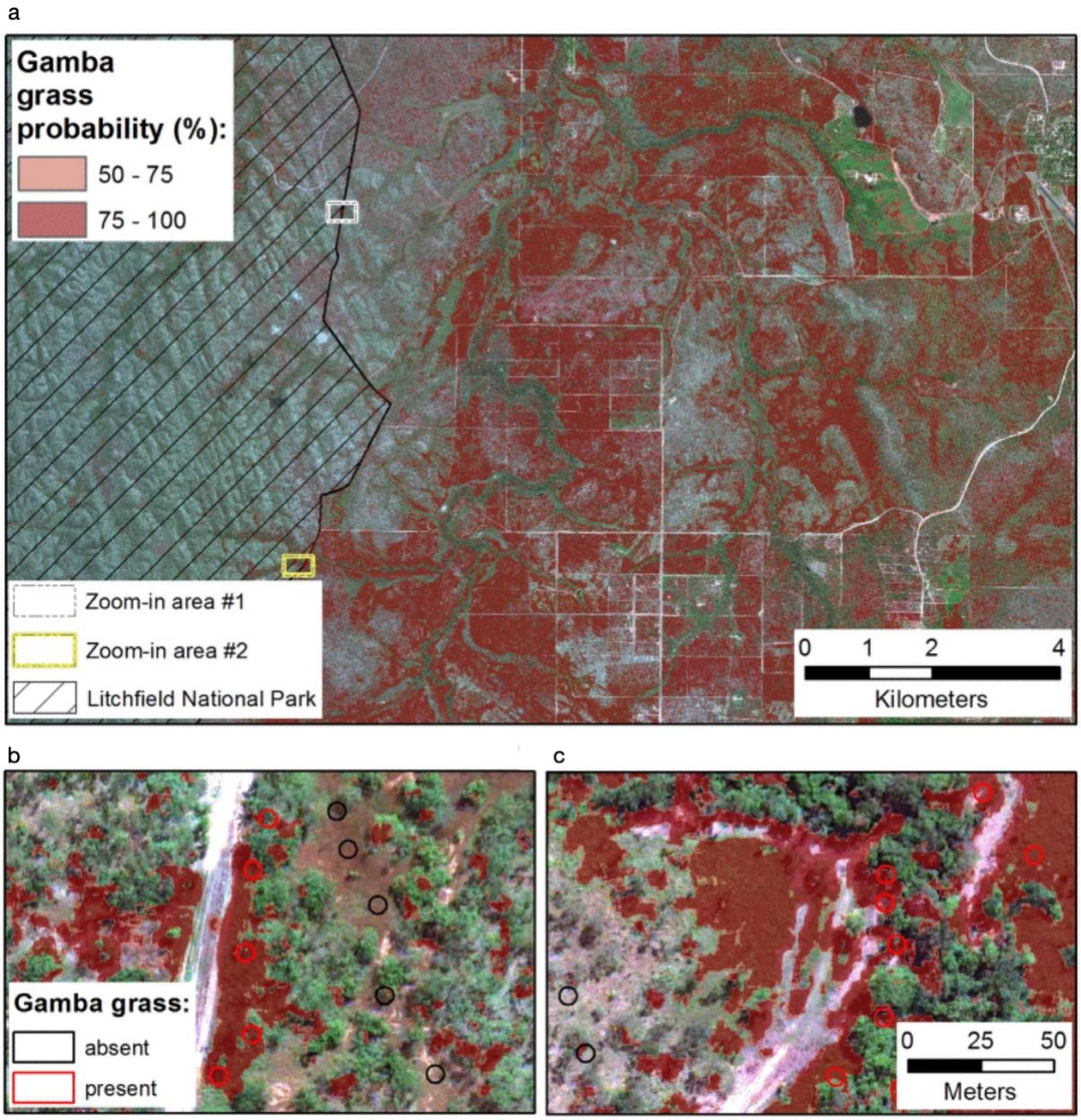


Figure 7. Spatial outputs from the final XGBoost model showing predictions of gamba grass occurrence at 0.3 m resolution for (a) the full study area and (b-c) zoomed-in examples.

4. Fusing open-access multi-spectral and synthetic aperture radar remote sensing for large-area gamba-grass mapping

The previous section illustrated how WorldView-3 imagery can be leveraged for mapping gamba grass presence at fine spatial resolutions with high accuracy. This outcome is an important step forward for intermediate-scale mapping, but the cost and logistics of WorldView-3 tasking and processing is prohibitive at regional scales. As such, in the second stage of our two-tiered approach, we used the high-resolution outputs from WorldView-3 to test the potential of open-access Sentinel-1 and Sentinel-2 imagery for cost-effective large-area mapping.

Initial investigations into the suitability of Sentinel-2 for this task showed that the reflectance data corresponded closely to patterns observed in the WorldView-3 data, albeit at a coarser resolution (Figure 8, Figure 9).

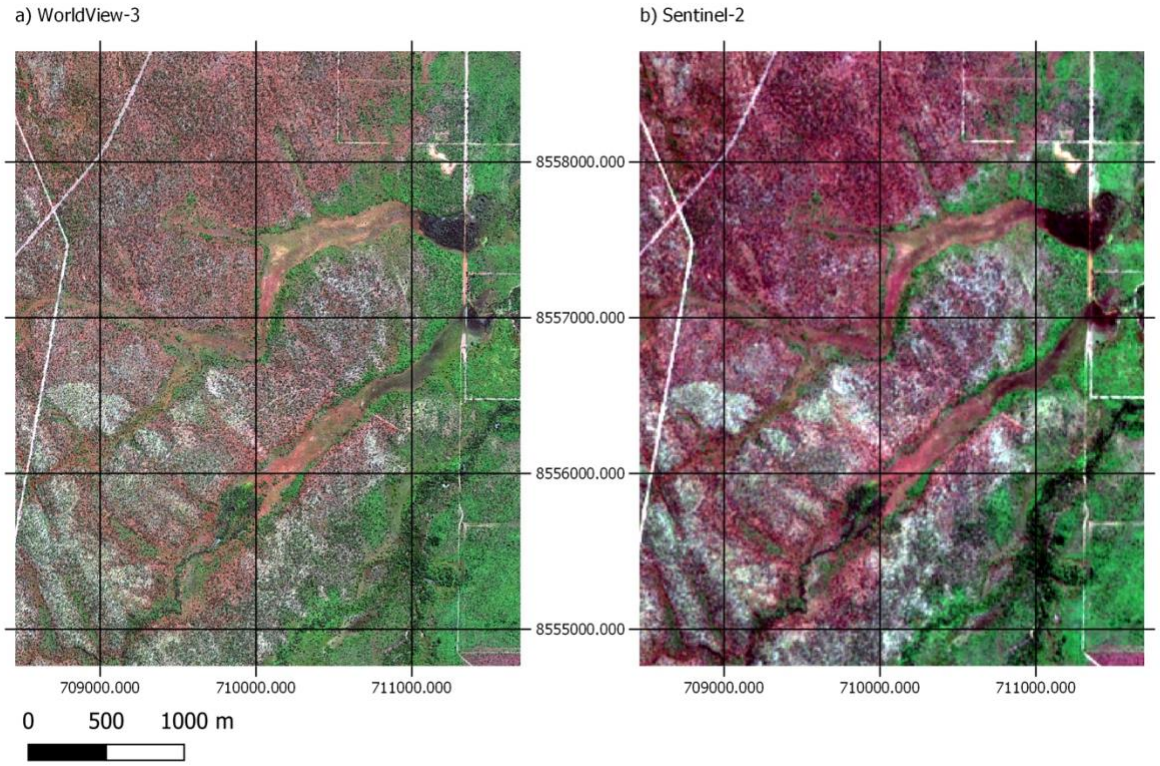


Figure 8. Comparison of (a) WorldView-3 and (b) Sentinel-2 imagery collected in April 2019. The Sentinel-2 scene was collected nine days later than the WorldView-3 image, but key patterns are well represented.

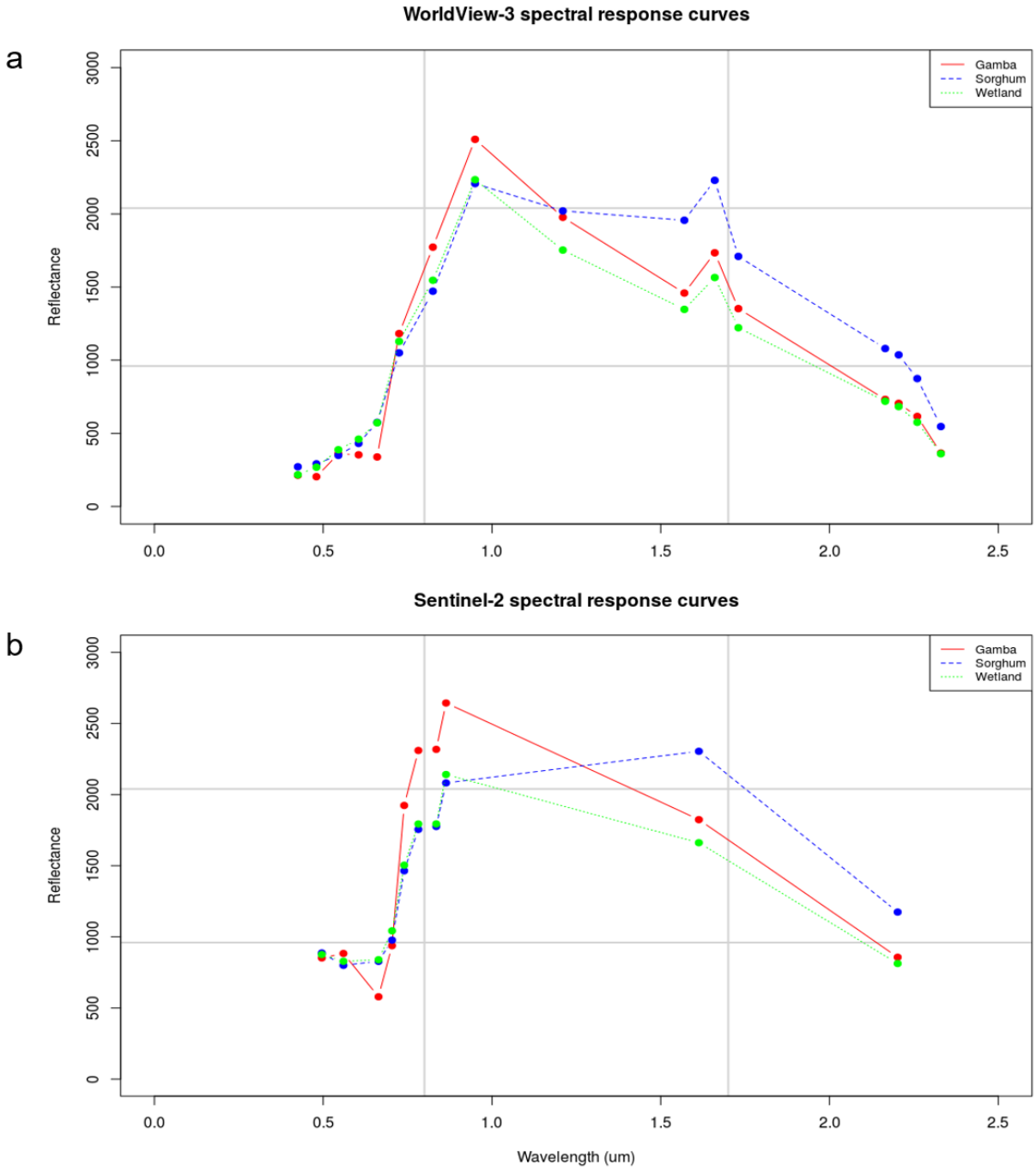


Figure 9. Spectral response curve of gamba grass, sorghum and native wetland grasses as recorded from (a) WorldView-3 and (b) Sentinel-2 satellites in April 2019. Open-access Sentinel-2 data captures key reflectance patterns very well despite having lower spatial and spectral resolution than WorldView-3.

4.1 Open-access imaging of vegetation structure and reflectance

Sentinel-1 and Sentinel-2 are the synthetic aperture radar (SAR) and multi-spectral (13-band) sister missions of the ESA's Copernicus program, providing repeat coverage over northern Australia every 6–10 days at 10–20 m resolution through their constellation orbit arrangement.

We used Open Data Cube infrastructure (CSIRO's EASI Hub and Digital Earth Australia's implementations) to query the Sentinel-1 and Sentinel-2 imagery intersecting a 5,000 km^2 area of interest that encompassed the WorldView-3 results from the previous section. We restricted the available image search to a 6-week period from 1 April to 15 May, to capture the peak in the phenological cycle for gamba grass, in 2019, 2020 and 2021. A maximum cloud cover threshold of 5% (per scene) was imposed on the Sentinel-2 search (surface reflectance product), but not on the Sentinel-1 search (ground-range detected product) since radar imagery is less impacted by cloud cover. A total of 21 suitable Sentinel-1 scenes and 57 Sentinel-2 scenes were extracted for analysis and stacked into a data-cube structure for model development and analytics.

We tested a suite of gradient-boosting regression trees using the scikit-learn machine-learning package in Python (Pedregosa et al. 2011). The full image series from the peak phenological season for 2019, 2020 and 2021 were included. Input variables consisted of the median raw reflectance data from Sentinel-2 (bands 1–12), spectral indices from Sentinel-2 (Normalised Difference Vegetation Index [NDVI], Enhanced Vegetation Index [EVI], Normalised Burn Ratio [NBR]), median backscatter intensities from Sentinel-1 (VV and VH polarisations), and terrain variables derived from SRTM imagery (elevation, slope). Sentinel-2 provides 10 m resolution in the VNIR region, but SWIR bands are at 20 m spatial resolution and we considered this more appropriate for large-area mapping. As such, the high-resolution gamba grass probability mapping from Section 3 (Figure 7) was resampled to 20 m spatial scale to provide input training data for the larger-area model (Figure 10). A total of 5,000 points were randomly distributed across the training dataset and the gamba grass probability was extracted for each point. These probabilities were reclassified into 25% increments from 0 to 100%, to provide the categorical inputs for the model. The gradient-boosted tree regression model was built with 500 trees, with a shrinkage parameter of 0.005. A sampling rate of 0.7 was specified, together with a maximum node allowance of 8. The Huber loss function was selected as it is less sensitive to outliers than squared error loss estimates.

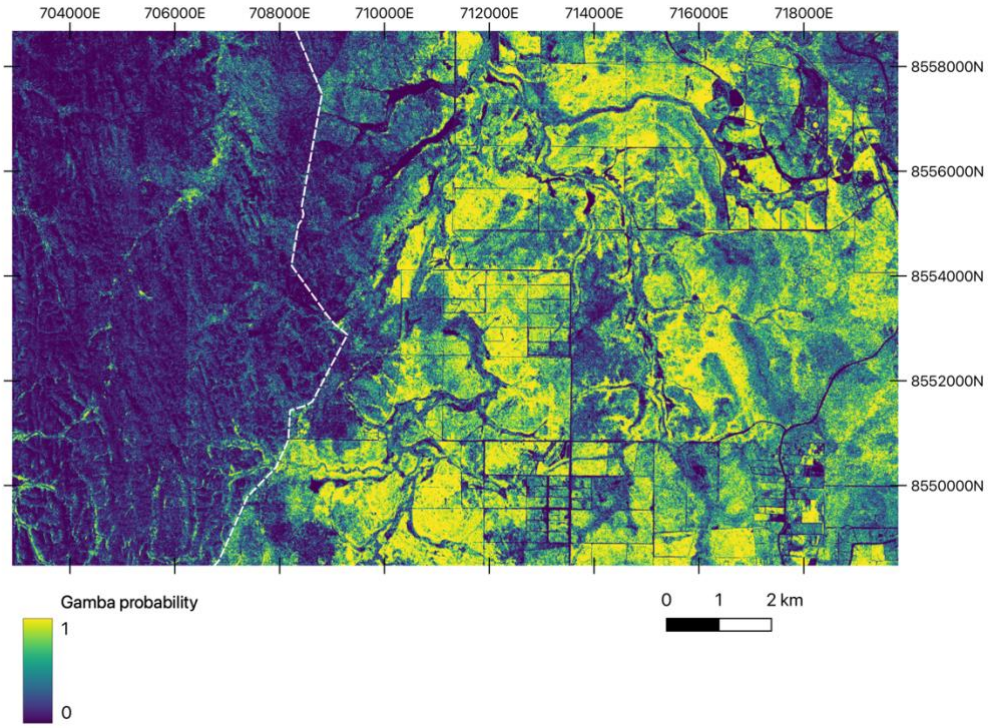


Figure 10. Training data generated from the WorldView-3 analysis (April 2019 data), aggregated to 20 m scale to match Sentinel-1/-2 inputs.

The red and SWIR1 wavelengths from Sentinel-2 and the VV backscatter polarisation from Sentinel-1 were the top three ranked variables. The resulting classification returned a BA of 82% for the gamba vs non-gamba split, and 68% for the separation into the four density categories. It is clear from the confusion matrix (Figure 11) that the reduction in overall accuracy when increasing the number of categories is a result of confusion between neighbouring classes: the high-density class showed some commission to the medium-density class, and the no-density class showed some commission to the low-density class.

Spatial outputs were well correlated with the locations of known gamba grass infestations but also identified some additional locations that require further investigation in the field. For example, the detection of infestations in north-east Litchfield National Park (Figure 12a,b) and southern Litchfield along the Reynold's River (Figure 12c,d) were in line with expectations, but the spatial extent of predictions in the Adelaide River hills (Figure 12e,f) needs further on-ground truthing.

While the native probability predictions from Sentinel-1/-2 are at 20 m resolution, they can be upscaled to any spatial unit of interest, such as 100 m (common ecological monitoring scale of 1 ha) and 250 m resolutions (reporting grid size under the ERF).

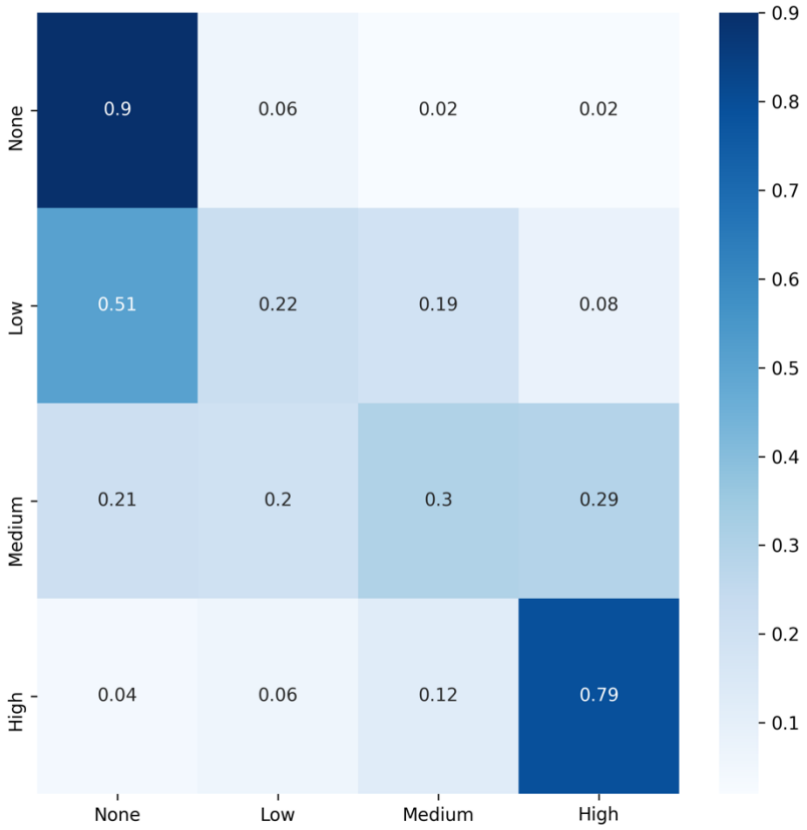


Figure 11. Confusion matrix for the classification of gamba grass occurrence into different density classes at the 20 m scale.

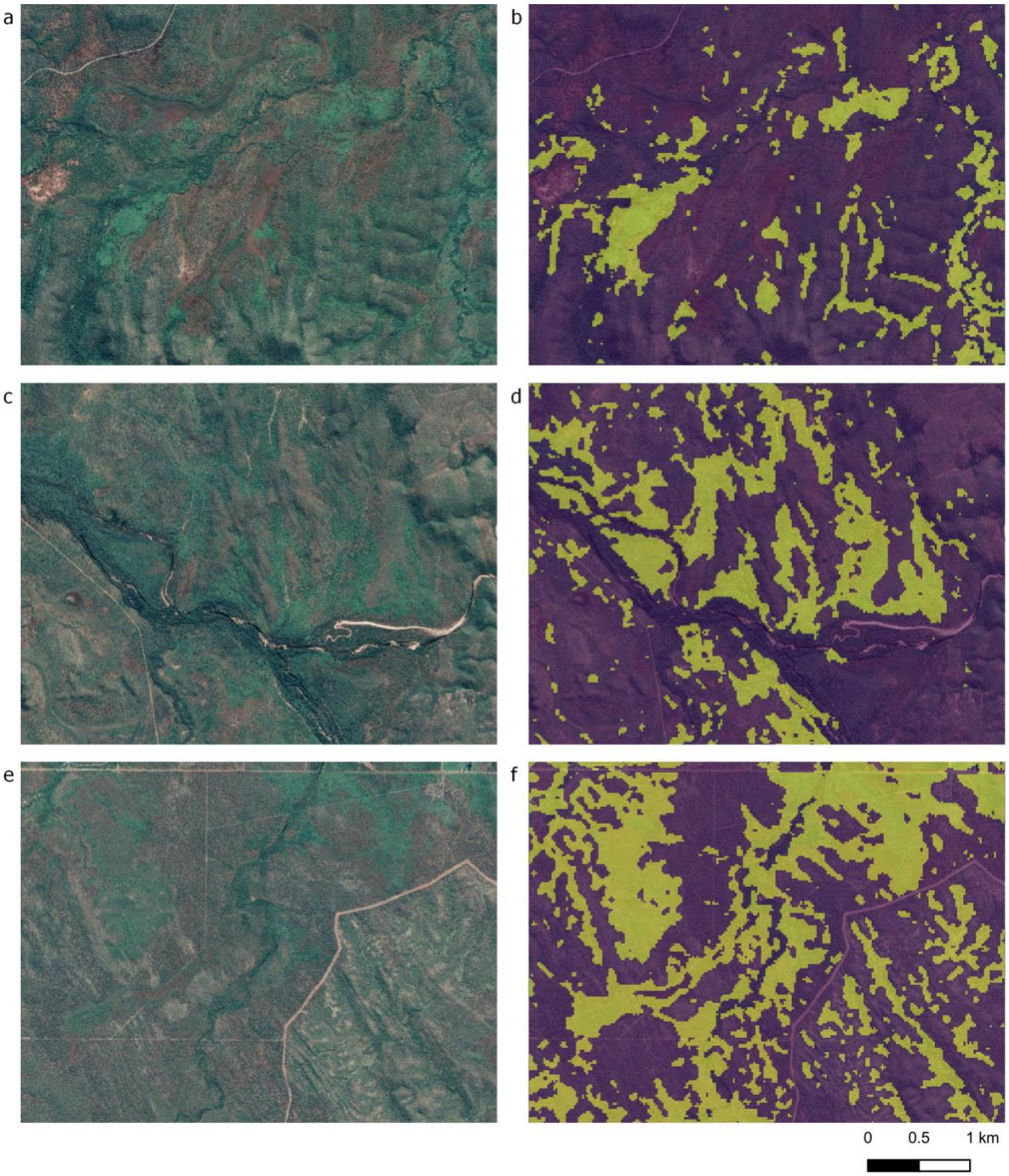


Figure 12. Predictions of gamba grass occurrence from Sentinel-1 and Sentinel-2 outside of the training area. Left column panels (a,c,e) are RGB composites from VHR satellite imagery with gamba grass visible in homogeneous green. Right column panels (b,d,f) show modelled results with gamba grass depicted in yellow and non-gamba in purple. a,b: north-east Litchfield National Park. c,d: Reynolds River. e,f: Adelaide River hills.

5. Accessibility, cost and processing requirements

A key factor in establishing a feasible remote sensing method for mapping and monitoring gamba grass on ERF properties is the accessibility and cost of the required data and processing requirements.

5.1 Accessibility and cost

The vast majority of current Earth observation programs from national agencies support free use and open access of imagery for research purposes. This trend was first established by the United States Geological Survey (USGS), who changed policy in 2008 to provide all Landsat data for free over the internet (Wulder et al. 2012). This policy shift was pivotal, and opening up the archive has had substantial downstream effects by growing the user base and leading to more efficient processing tools and analysis algorithms (Turner et al. 2015). Landsat 8 data remains freely available (Table 1).

The ESA's Copernicus program has adopted a similar model in making all Sentinel products accessible to the public (Table 1). This is a significant factor in prioritising their use in gamba grass monitoring for the ERF.

Open access to national space agency imagery is commendable, but these agencies cannot cover all bases and niche markets exist for commercial providers to meet specialist requirements. The primary commercial satellite used in this study is Digital Globe's WorldView-3. The tasking of this satellite currently costs approximately A$4,000 for a minimum order size of 100 km^2. Similar specification (multi-spectral, 0.5 m spatial resolution) options can also be acquired from KOMPSAT-3 and Pleiades for A$2,000–A$3,000 per 100 km^2 depending on the timeliness of the dataset (Table 1). These VHR commercial satellites have the additional advantage of flexible tasking times to align with particular periods of the year that are of phenological interest. Given the property-scale size of 300–30,000 km^2, these satellites are cost prohibitive for large-area monitoring, but the accuracy they provide is critical for building large calibration and validation libraries.

5.2 Data cubes and cloud-based distributed processing

A further consideration for mapping and monitoring of gamba grass for the ERF methodology is the data storage and processing requirements. Local storage and processing of imagery is inefficient and unfeasible for mapping tasks of large spatial extent and with rich time-series data (Lewis et al. 2017). Given the huge volumes of satellite data collected daily, and the large spatial area of properties in northern Australia, a mapping campaign aimed at characterising high-biomass grasses will need to use high-performance storage and computing clusters. Fortunately, computing power and access to servers is no longer a restriction or a major limiting expense (Wulder et al. 2018). Cloud-based distributed storage and processing platforms like the Open Data Cube and Google Earth Engine have made access and processing of image archives accessible, transparent and standardised (Mueller et al. 2016, Dhu et al. 2017)Gorelick et al. 2017). Within the Australian context, the development of Digital Earth Australia and the CSIRO's EASI Hub is facilitating the analysis of continental-scale time-series of satellite data through an implementation of the Open Data

Cube, and implements geometric and spectral corrections for comparability of measurements across the time-series (Lewis et al. 2017). These data storage and computer infrastructures make complex time-series analysis accessible and feasible over large areas, and could be used more heavily in future efforts to increase the precision of gamba-grass mapping, and in particular, to begin tracking changes over time (Figure 13).

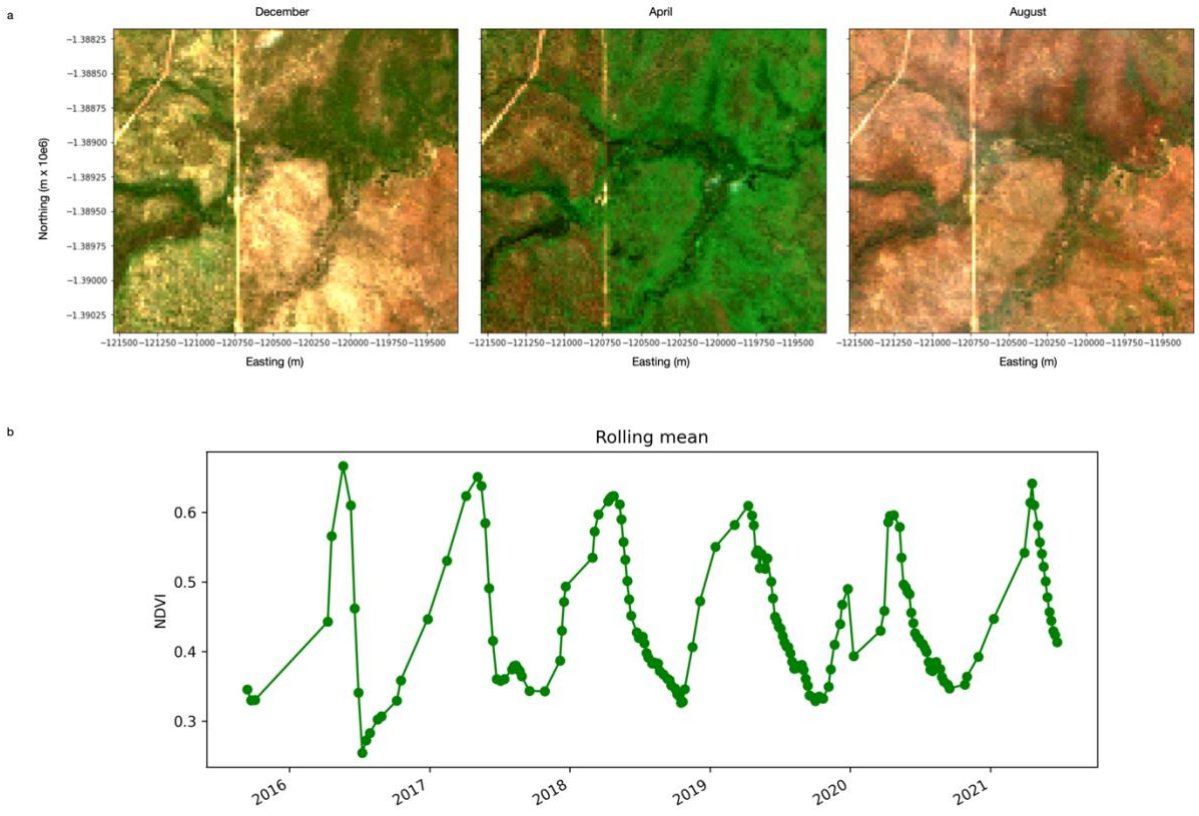


Figure 13. The high temporal resolution of Sentinel-2 enables in-depth assessment of vegetation phenology over time (Litchfield–Batchelor boundary). Image extracts based on Jupyter Notebooks developed by Digital Earth Australia (Krause 2021).

5.3 Future developments in Earth observation

Earth observation from space is a rapidly expanding frontier, and multiple space agencies around the globe have new satellites in development for launch in the coming years. There are numerous missions on the horizon that could be of benefit to gamba-grass mapping, and these should be kept in mind when developing the gamba-grass mapping and monitoring strategy for ERF purposes.

Notable among the upcoming missions from an invasive species monitoring perspective is the planned launch of three major spaceborne hyperspectral missions within the next decade – EnMAP from the German Aerospace Centre (DLR), CHIME from the ESA, and SBG from NASA. These missions will operate globally at the 30 m scale, providing rich spectral detail across the VNIR and SWIR portions of the electromagnetic spectrum.

Insights into how these systems might perform can already be gained through the assessment of experimental imagery captured by the DLR's DESIS, a VNIR hyperspectral imager onboard the International Space Station. A number of DESIS scenes have been collected over northern Australia already, and an example is shown in Figure 14. As opposed to sampling a handful of wavelengths, DESIS captures 232 bands within the VNIR portion of the spectrum and provides richer detail on pigment differences in leaf material. Despite the image shown in Figure 14 being collected in the middle of the dry season, clear distinction between gamba grass and native riparian grasses is evident in the signatures both above and below the 700 nm inflection point.

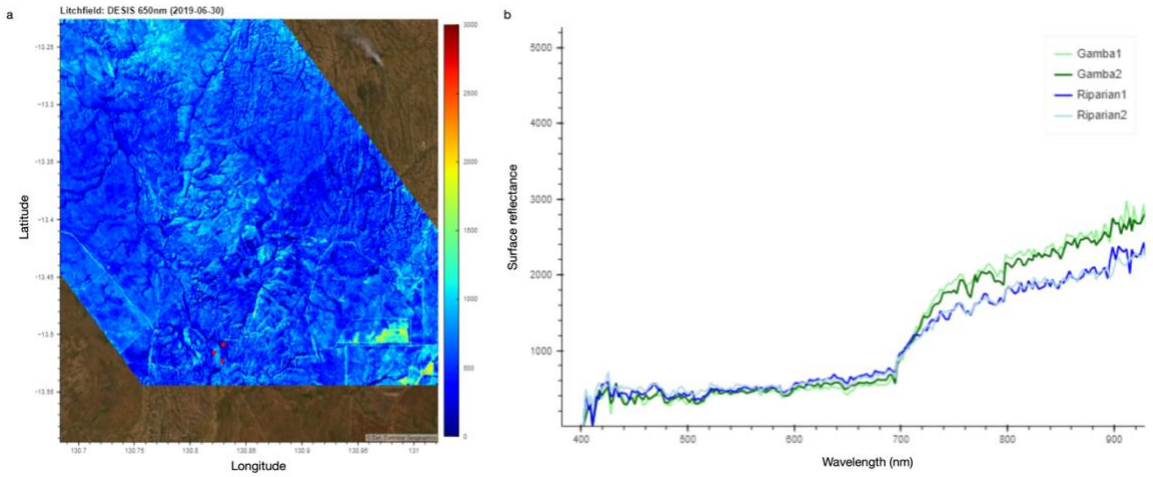


Figure 14. Spectral reflectance curve derived from the experimental DESIS hyperspectral mission. High spectral resolution measurements from space will enable finer delineation between species based on pigment and leaf structure attributes.

From a structural perspective, the launch of NISAR in 2023 will provide high temporal frequency L-band coverage at 10 m resolution. NISAR is a joint NASA/ISRO mission, and the longer wavelength of the L-band sensor (as opposed to the C-band of Sentinel-1) will be well-suited to biomass mapping applications and could help inform the mapping of dense gamba grass stands.

The next decade will therefore continue to be very exciting from an Earth observation perspective. However, irrespective of the developments that take place in terms of sensor and satellite technologies, reliable mapping and quantification of accuracies and confidence limits will ultimately hinge on robust calibration and validation data from the field and/or from proximal remote sensing. As such, the reference endmembers and classification results produced in the next few years will hold great value for calibration of future space missions.

6. Guidelines for gamba-grass mapping with remote sensing in the ERF context

The combination of Sentinel-1 and Sentinel-2 data streams presents a compelling avenue for large-area mapping at ecologically useful spatial (10–20 m) and temporal resolutions (6–10 days) across northern Australia. The fusion of optical reflectance and radar backscatter data provides a rich data cube of structural and spectral properties that is freely available and readily accessible from cloud-based distributed computing platforms. A large training dataset (5,000 points) was essential for building machine-learning models to map gamba grass presence at large spatial scales (5,000 km^2). Obtaining the required volume of training data is not possible from fieldwork alone, but the two-tiered approach developed here illustrates how very-high resolution imagery (e.g. WorldView-3/KOMPSAT-3/Pleiades/Planet/unmanned aerial vehicle [UAV]) can be leveraged for this purpose. Gradient-boosted tree regression performed well on both the high-resolution mapping with WorldView-3 imagery and on the large-area composite images produced from Sentinel-1/-2. More complex machine-learning approaches that account for temporal context within the time-series more explicitly (e.g. recurrent neural networks) could help refine results and improve precision, but will add to the computational complexity. The core elements from this workflow could be implemented on properties of interest in multiple ways, either using open-access imagery only or leveraging VHR satellite/UAV imagery if available (Figure 15).

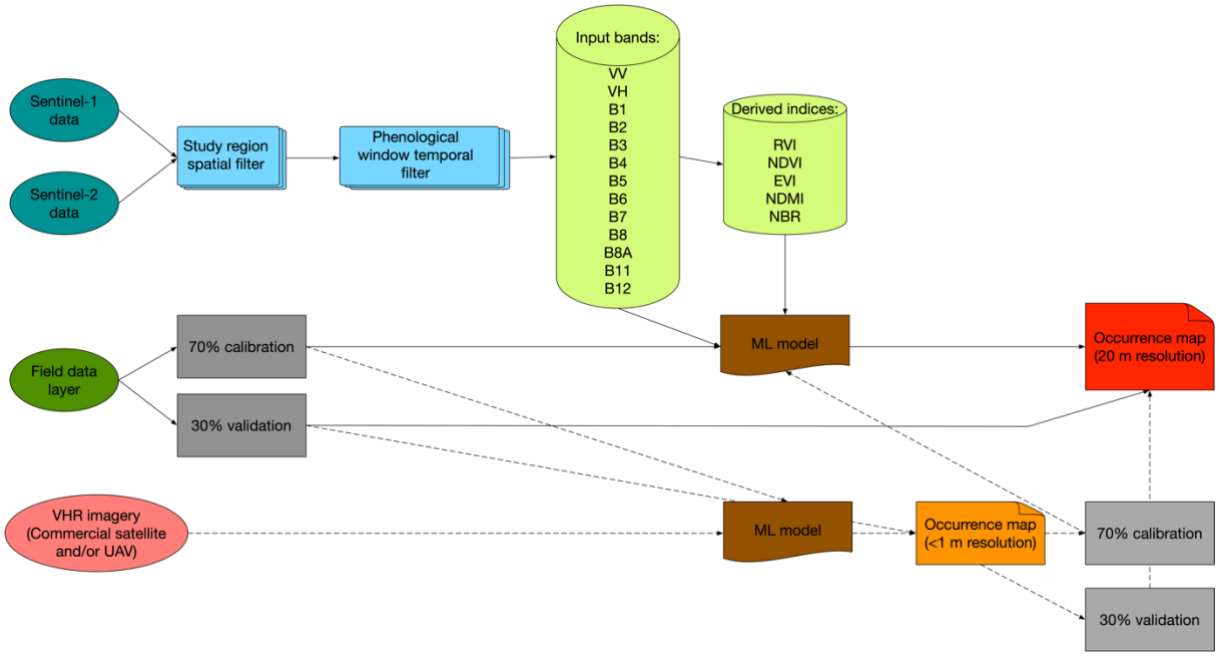


Figure 15. Suggested workflow for mapping gamba grass on Emissions Reduction Fund properties. Solid lines represent the fully open-access workflow which would be widely usable/applicable. The dashed lines indicate additional steps required for the hierarchical approach used in this study, leveraging commercial satellite imagery (and/or imaging from unmanned aerial vehicles). The machine-learning (ML) model can vary in complexity from standard statistical relationships to more complex deep learning models depending on the user skillset and available processing resources.

6.1.1 Guideline considerations

Ecological

- Phenological timing is critical for the success of gamba-grass mapping activities with remotely sensed imagery.
- Optimal phenological windows for image acquisition will vary across northern Australia and need to be adjusted regionally.
- Optimal phenological windows can only be broadly predicted by date; rainfall in the preceding months is very influential.

Technical – remote sensing

- VHR satellite imagery and/or aerial/UAV imagery can provide an important steppingstone in the scaling chain.
- Field-collected samples will often be insufficient for the training of models running at large regional scales from satellite imagery.
- Local-scale mapping results from VHR imagery should be nested within medium-resolution satellite imagery to provide a scaling mechanism from field data to large-area satellite data.

Technical – machine learning

- Robust training and validation libraries are critical for the development of reliable models.
- Machine-learning models require very large datasets for training and validation purposes.
- In addition to gamba and non-gamba classes, it is important to include the presence of similar native species (particularly high-biomass wetland grasses) in the training and validation libraries.

References

Adams, V. M., and S. A. Setterfield. 2013. Estimating the financial risks of Andropogon gayanus to greenhouse gas abatement projects in northern Australia. Environmental Research Letters **8**.

Ali, I., F. Cawkwell, E. Dwyer, B. Barrett, and S. Green. 2016. Satellite remote sensing of grasslands: from observation to management. Journal of Plant Ecology **9**:649-671.

Alvarez-Taboada, F., C. Paredes, and J. Julián-Pelaz. 2017. Mapping of the invasive species Hakea sericea using unmanned aerial vehicle (UAV) and WorldView-2 imagery and an object-oriented approach. Remote Sensing **9**:913.

Baraldi, A., and F. Parmiggiani. 1995. An Investigation of the Textural Characteristics Associated with Gray-Level Cooccurrence Matrix Statistical Parameters. IEEE Transactions on Geoscience and Remote Sensing **33**:293-304.

Belward, A. S., and J. O. Skoien. 2015. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. Isprs Journal of Photogrammetry and Remote Sensing **103**:115-128.

Bergstra, J., N. Pinto, and D. Cox. 2012. Machine learning for predictive auto-tuning with boosted regression trees. Pages 1-9 *in* 2012 Innovative Parallel Computing (InPar). IEEE.

Beringer, J., L. B. Hutley, D. Abramson, S. K. Arndt, P. Briggs, M. Bristow, J. G. Canadell, L. A. Cernusak, D. Eamus, A. C. Edwards, B. J. Evans, B. Fest, K. Goergen, S. P. Grover, J. Hacker, V. Haverd, K. Kanniah, S. J. Livesley, A. Lynch, S. Maier, C. Moore, M. Raupach, J. Russell-Smith, S. Scheiter, N. J. Tapper, and P. Uotila. 2015. Fire in Australian savannas: from leaf to landscape. Global Change Biology **21**:62-81.

Blaschke, T., G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Q. Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede. 2014. Geographic Object-Based Image Analysis - Towards a new paradigm. Isprs Journal of Photogrammetry and Remote Sensing **87**:180-191.

BOM. 2020. Climate Statistics for Australian Locations.

Bradley, B. A. 2014. Remote detection of invasive plants: a review of spectral, textural and phenological approaches. Biological Invasions **16**:1411-1425.

Briem, G. J., J. A. Benediktsson, and J. R. Sveinsson. 2002. Multiple classifiers applied to multisource remote sensing data. IEEE Transactions on Geoscience and Remote Sensing **40**:2291-2299.

Brooks, K. J., S. A. Setterfield, and M. M. Douglas. 2010. Exotic Grass Invasions: Applying a Conceptual Framework to the Dynamics of Degradation and Restoration in Australia's Tropical Savannas. Restoration Ecology **18**:188-197.

Chen, T. Q., and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. Kdd'16: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining:785-794.

Clerici, N., C. A. V. Calderon, and J. M. Posada. 2017. Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: a case study in the lower Magdalena region, Colombia. Journal of Maps **13**:718-726.

Dhu, T., B. Dunn, B. Lewis, L. Lymburner, N. Mueller, E. Telfer, A. Lewis, A. McIntyre, S. Minchin, and C. Phillips. 2017. Digital earth Australia – unlocking new value from earth observation data. Big Earth Data **1**:64-74.

Donlon, C., B. Berruti, A. Buongiorno, M. H. Ferreira, P. Femenias, J. Frerick, P. Goryl, U. Klein, H. Laur, C. Mavrocordatos, J. Nieke, H. Rebhan, B. Seitz, J. Stroede, and R. Sciarra. 2012. The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. Remote Sensing of Environment **120**:37-57.

Drusch, M., U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini. 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment **120**:25-36.

Ens, E., L. B. Hutley, N. A. Rossiter-Rachor, M. M. Douglas, and S. A. Setterfield. 2015. Resource-use efficiency explains grassy weed invasion in a low-resource savanna in north Australia. Frontiers in Plant Science **6**.

Evangelista, P. H., T. J. Stohlgren, J. T. Morisette, and S. Kumar. 2009. Mapping Invasive Tamarisk (Tamarix): A Comparison of Single-Scene and Time-Series Analyses of Remotely Sensed Data. Remote Sensing **1**:519-533.

Ferdinands, K. B., M. M. Douglas, S. A. Setterfield, and J. L. Barratt. 2006. Africanising the tropical woodlands: Canopy loss and tree death following gamba grass Andropogon gayanus invasion. Pages 296-296 *in* 15th Australian Weeds Conference: Managing Weeds in a Changing Climate. Weed Management Society of South Australia.

Gavier-Pizarro, G. I., T. Kuemmerle, L. E. Hoyos, S. I. Stewart, C. D. Huebner, N. S. Keuler, and V. C. Radeloff. 2012. Monitoring the invasion of an exotic tree (Ligustrum lucidum) from 1983 to 2006 with Landsat TM/ETM plus satellite data and Support Vector Machines in Cordoba, Argentina. Remote Sensing of Environment **122**:134-145.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment **202**:18-27.

Haldar, S. K. 2013. Mineral Exploration: Principles and Applications. Mineral Exploration: Principles and Applications:1-334.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science **342**:850-853.

Hill, M. J., U. Senarath, A. Lee, M. Zeppel, J. M. Nightingale, R. D. J. Williams, and T. R. McVicar. 2006. Assessment of the MODIS LAI product for Australian ecosystems. Remote Sensing of Environment **101**:495-518.

Jordan, M. I., and T. M. Mitchell. 2015. Machine learning: Trends, perspectives, and prospects. Science **349**:255-260.

Krause, C., Dunn, B., Bishop-Taylor, R., Adams, C., Burton, C., Alger, M., Chua, S., Phillips, C., Newey, V., Kouzoubov, K., Leith, A., Ayers, D., Hicks, A., DEA Notebooks contributors 2021. 2021. Digital Earth Australia notebooks and tools repository. Geoscience Australia, Canberra.

Lass, L. W., T. S. Prather, N. F. Glenn, K. T. Weber, J. T. Mundt, and J. Pettingill. 2005. A review of remote sensing of invasive weeds and example of the early detection of

spotted knapweed (Centaurea maculosa) and babysbreath (Gypsophila paniculata) with a hyperspectral sensor. Weed Science **53**:242-251.

Lawrence, R. L., S. D. Wood, and R. L. Sheley. 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). Remote Sensing of Environment **100**:356-362.

Levick, S. R., and K. H. Rogers. 2011. Context-dependent vegetation dynamics in an African savanna. Landscape Ecology **26**:515-528.

Lewis, A., S. Oliver, L. Lymburner, B. Evans, L. Wyborn, N. Mueller, G. Raevksi, J. Hooke, R. Woodcock, J. Sixsmith, W. J. Wu, P. Tan, F. Q. Li, B. Killough, S. Minchin, D. Roberts, D. Ayers, B. Bala, J. Dwyer, A. Dekker, T. Dhu, A. Hicks, A. Ip, M. Purss, C. Richards, S. Sagar, C. Trenham, P. Wang, and L. W. Wang. 2017. The Australian Geoscience Data Cube - Foundations and lessons learned. Remote Sensing of Environment **202**:276-292.

Li, J., and D. P. Roy. 2017. A Global Analysis of Sentinel-2A, Sentinel-2B and Landsat-8 Data Revisit Intervals and Implications for Terrestrial Monitoring. Remote Sensing **9**.

Maier, S. W. 2010. Changes in surface reflectance from wildfires on the Australian continent measured by MODIS. International Journal of Remote Sensing **31**:3161-3176.

Main, R., R. Mathieu, W. Kleynhans, K. Wessels, L. Naidoo, and G. P. Asner. 2016. Hyper-Temporal C-Band SAR for Baseline Woody Structural Assessments in Deciduous Savannas. Remote Sensing **8**.

Marshall, V., M. Lewis, and B. Ostendorf. 2012. Do Additional Bands (Coastal, Nir-2, Red-Edge and Yellow) in Worldview-2 Multispectral Imagery Improve Discrimination of an Invasive Tussock, Buffel Grass (Cenchrus Ciliaris)? Xxii Isprs Congress, Technical Commission Viii **39-B8**:277-281.

Matongera, T. N., O. Mutanga, T. Dube, and M. Sibanda. 2017. Detection and mapping the spatial distribution of bracken fern weeds using the Landsat 8 OLI new generation sensor. International Journal of Applied Earth Observation and Geoinformation **57**:93-103.

Miller, A. 2002. Subset selection in regression. CRC Press.

Mueller, N., A. Lewis, D. Roberts, S. Ring, R. Melrose, J. Sixsmith, L. Lymburner, A. McIntyre, P. Tan, S. Curnow, and A. Ip. 2016. Water observations from space: mapping surface water from 25 years of Landsat imagery across Australia. Remote Sensing of Environment **174**:341–352.

Olsony, R. S., W. La Cava, Z. Mustahsan, A. Varik, and J. H. Moore. 2018. Data-driven advice for applying machine learning to bioinformatics problems. Pacific Symposium on Biocomputing 2018 (Psb):192-203.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research **12**:2825-2830.

Pettorelli, N., J. O. Vik, A. Mysterud, J. M. Gaillard, C. J. Tucker, and N. C. Stenseth. 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in Ecology & Evolution **20**:503-510.

Petty, A. M., S. A. Setterfield, K. B. Ferdinands, and P. Barrow. 2012. Inferring habitat suitability and spread patterns from large-scale distributions of an exotic invasive pasture grass in north Australia. Journal of Applied Ecology **49**:742-752.

Pflugmacher, D., W. B. Cohen, and R. E. Kennedy. 2012. Using Landsat-derived disturbance history (1972-2010) to predict current forest structure. Remote Sensing of Environment **122**:146-165.

Robinson, T. P., G. W. Wardell-Johnson, G. Pracilio, C. Brown, R. Corner, and R. D. van Klinken. 2016. Testing the discrimination and detection limits of WorldView-2 imagery on a challenging invasive plant target. International Journal of Applied Earth Observation and Geoinformation **44**:23-30.

Rossiter-Rachor, N. A., S. A. Setterfield, M. M. Douglas, L. B. Hutley, and G. D. Cook. 2008. Andropogon gayanus (Gamba grass) invasion increases fire-mediated nitrogen losses in the tropical savannas of northern Australia. Ecosystems **11**:77-88.

Rossiter-Rachor, N. A., S. A. Setterfield, M. M. Douglas, L. B. Hutley, G. D. Cook, and S. Schmidt. 2009. Invasive Andropogon gayanus (gamba grass) is an ecosystem transformer of nitrogen relations in Australian savanna. Ecological Applications **19**:1546-1560.

Rossiter, N. A., S. A. Setterfield, M. M. Douglas, and L. B. Hutley. 2003. Testing the grass-fire cycle: alien grass invasion in the tropical savannas of northern Australia. Diversity and Distributions **9**:169-176.

Schmidt, J., F. E. Fassnacht, M. Forster, and S. Schmidtlein. 2018. Synergetic use of Sentinel-1 and Sentinel-2 for assessments of heathland conservation status. Remote Sensing in Ecology and Conservation **4**:225-239.

Setterfield, S., M. Douglas, L. Hutley, K. Ferdinands, E. Ens, K. Brooks, and N. Rossiter. 2008. Ecosystem impacts of an exotic grass in northern Australia: effects on structure and carbon stocks. Page 198 *in* Sixteenth Australian Weeds Conference. Queensland Weeds Society, Brisbane.

Setterfield, S. A., N. A. Rossiter-Rachor, and V. M. Adams. 2018. Navigating the fiery debate: the role of scientific evidence in eliciting policy and management responses for contentious plants in northern Australia. Pacific Conservation Biology **24**:318-328.

Setterfield, S. A., N. A. Rossiter-Rachor, M. M. Douglas, L. Wainger, A. M. Petty, P. Barrow, I. J. Shepherd, and K. B. Ferdinands. 2013. Adding Fuel to the Fire: The Impacts of Non-Native Grass Invasion on Fire Management at a Regional Scale. Plos One **8**.

Setterfield, S. A., N. A. Rossiter-Rachor, L. B. Hutley, M. M. Douglas, and R. J. Williams. 2010. Turning up the heat: the impacts of Andropogon gayanus (gamba grass) invasion on fire behaviour in northern Australian savannas. Diversity and Distributions **16**:854-861.

Skidmore, A. K., J. G. Ferwerda, O. Mutanga, S. E. Van Wieren, M. Peel, R. C. Grant, H. H. T. Prins, F. B. Balcik, and V. Venus. 2010. Forage quality of savannas - Simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. Remote Sensing of Environment **114**:64-72.

Torres, R., P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, I. N. Traver, P. Deghaye, B. Duesmann, B. Rosich, N. Miranda, C. Bruno, M. L'Abbate, R. Croci, A. Pietropaolo, M. Huchler, and F. Rostan. 2012. GMES Sentinel-1 mission. Remote Sensing of Environment **120**:9-24.

Turner, W., C. Rondinini, N. Pettorelli, B. Mora, A. K. Leidner, Z. Szantoi, G. Buchanan, S. Dech, J. Dwyer, M. Herold, L. P. Koh, P. Leimgruber, H. Taubenboeck, M. Wegmann, M. Wikelski, and C. Woodcock. 2015. Free and open-access satellite data are key to biodiversity conservation. Biological Conservation **182**:173-176.

Turner, W., S. Spector, N. Gardiner, M. Fladeland, E. Sterling, and M. Steininger. 2003. Remote sensing for biodiversity science and conservation. Trends in Ecology & Evolution **18**:306-314.

Whiteside, T. G., G. S. Boggs, and S. W. Maier. 2011. Comparing object-based and pixel-based classifications for mapping savannas. International Journal of Applied Earth Observation and Geoinformation **13**:884-893.

Wood, E. M., A. M. Pidgeon, V. C. Radeloff, and N. S. Keuler. 2012. Image texture as a remotely sensed measure of vegetation structure. Remote Sensing of Environment **121**:516-526.

Wooster, M. J., W. Xu, and T. Nightingale. 2012. Sentinel-3 SLSTR active fire detection and FRP product: Pre-launch algorithm development and performance evaluation using MODIS and ASTER datasets. Remote Sensing of Environment **120**:236-254.

Wulder, M. A., N. C. Coops, D. P. Roy, J. C. White, and T. Hermosilla. 2018. Land cover 2.0. International Journal of Remote Sensing **39**:4254-4284.

Wulder, M. A., J. G. Masek, W. B. Cohen, T. R. Loveland, and C. E. Woodcock. 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. Remote Sensing of Environment **122**:2-10.

Xiong, X., J. Butler, K. Chiang, B. Efremova, J. Fulbright, N. Lei, J. McIntire, H. Oudrari, J. Sun, Z. Wang, and A. Wu. 2014. VIIRS on-orbit calibration methodology and performance. Journal of Geophysical Research: Atmospheres **119**:5065-5078.

Yang, C. H., and J. H. Everitt. 2010. Mapping three invasive weeds using airborne hyperspectral imagery. Ecological Informatics **5**:429-439.

Zhang, Y. 2002. Problems in the fusion of commercial high-resolution satelitte as well as Landsat 7 images and initial solutions. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences **34**:587-592.

Zhou, J. J., R. Y. Guo, M. T. Sun, T. J. G. L. T. Di, S. Wang, J. Y. Zhai, and Z. Zhao. 2017. The Effects of GLCM parameters on LAI estimation using texture values from Quickbird Satellite Imagery. Scientific Reports **7**.